

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Space Programs Summary 37-48, Vol. III

Supporting Research and Advanced Development

For the Period October 1 to November 30, 1967

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

December 31, 1967

SPACE PROGRAMS SUMMARY 37-48, VOL. III

Copyright © 1968

Jet Propulsion Laboratory
California Institute of Technology

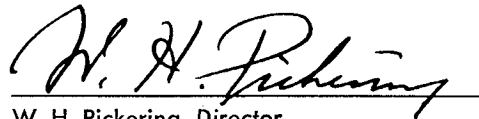
Prepared Under Contract No. NAS 7-100
National Aeronautics & Space Administration

Preface

The Space Programs Summary is a bimonthly publication that presents a review of engineering and scientific work performed, or managed, by the Jet Propulsion Laboratory for the National Aeronautics and Space Administration during a two-month period. Beginning with the 37-47 series, the Space Programs Summary is composed of four volumes:

- Vol. I. *Flight Projects* (Unclassified)
- Vol. II. *The Deep Space Network* (Unclassified)
- Vol. III. *Supporting Research and Advanced Development* (Unclassified)
- Vol. IV. *Flight Projects and Supporting Research and Advanced Development* (Confidential)

Approved by:

A handwritten signature in dark ink, appearing to read "W. H. Pickering", is written over a horizontal line.

W. H. Pickering, Director
Jet Propulsion Laboratory

Contents

SYSTEMS DIVISION

I. Systems Analysis Research	1
A. On the Statistics of Particle Fluxes Resulting From Solar Flares	
NASA Code 120-26-07-01, P. Wesseling	1
B. An Alternate Expression for Light Time Using General Relativity	
NASA Code 129-04-04-02, D. Holdridge	2
C. A Simple Approach to Gravitational Theory	
NASA Code 129-04-01-01, H. Lass	4
D. Introduction of New Planetary Masses Into Ephemeris Development Computations	
NASA Code 129-04-01-01, J. D. Mulholland	6
E. Collection of Optical and Radar-Range Data on the Planets	
NASA Code 129-04-04-02, D. A. O'Handley	7
F. On a New Necessary Condition for Optimal Control Problems With Discontinuities	
NASA Code 125-17-04-05, P. Dyer and S. R. McReynolds	9
II. Systems Analysis	13
A. Analysis of De-orbit Maneuver Execution Errors	
NASA Code 684-30-01-10, E. G. Piaggi	13
B. Analytical Determination of Midcourse Velocity Probability Distributions	
NASA Code 684-30-01-10, L. Kingsland	20
III. Computation and Analysis	26
A. Abstracts of Certain Mathematical Subroutines	
NASA Code 129-04-04-01, R. Hanson	26
B. A Numerical Integration of Lunar Motion Employing a Consistent Set of Constants	
NASA Code 129-04-04-01, C. J. Devine	33
C. Systems of Ordinary Differential Equations With an Algebraic Constraint	
NASA Code 129-04-04-01, A. J. Semtner	39
D. Reduction of a Problem in Relativistic Cosmology Using a Transformation Algorithm	
NASA Code 129-04-04-01, A. J. Semtner	42

PROJECT ENGINEERING DIVISION

IV. System Design and Integration	45
A. Advanced Development at the System Level	
NASA Code 186-68-09-09, K. Casani and J. Gardner	45

Contents (contd)

GUIDANCE AND CONTROL DIVISION

V. Spacecraft Power	49
A. Sterilizable Battery	
NASA Code 120-34-01-03, 05, 06, 10, 11, 12, 13, and 18, R. Lutwack	49
B. Solar Cell Standardization	
NASA Code 120-33-01-03, R. F. Greenwood	50
C. Feasibility Study: 30-W/lb Roll-up Solar Array	
NASA Code 120-33-01-07, W. A. Hasbach	51
D. Active Electronic Load	
NASA Code 120-33-02-01, G. Stapfer	57
E. Thermionic Development	
NASA Code 120-33-02-01 and 120-27-06-07, P. Rouklove	58
VI. Guidance and Control Analysis and Integration	64
A. Capsule System Advanced Development Operational Support Equipment	
NASA Code 120-33-08-08, K. Mussen	64
B. Gas Valve Flow Detector	
NASA Code 186-68-02-27, S. D. Moss	66
VII. Guidance and Control Research	70
A. Preparation and Physical Properties of α -Se	
NASA Code 129-02-05-09, S. Iizima and M-A. Nicolet	70
B. Magneto-Optic Information Storage	
NASA Code 129-02-05-06, G. Lewicki	72
C. Apparent Work Function of Cavity Emitters	
NASA Code 129-02-01-07, K. Shimada	73

ENGINEERING MECHANICS DIVISION

VIII. Materials	77
A. Nonmagnetic Interconnect Material for Welded Modules	
NASA Code 186-68-10-09, R. E. Ringsmuth	77
B. Planetary Entry Heat Shields	
NASA Code 124-08-03-02, T. F. Moran	78
IX. Electronics Parts Engineering	82
A. Accelerated Testing Concepts, Methodology, and Models: A Literature Review	
NASA Code 186-70-01-05, E. Klippenstein	82

Contents (contd)

PROPULSION DIVISION

X. Solid Propellant Engineering	85
A. Applications Technology Satellite Motor Development NASA Code 630-01-00-00, R. G. Anderson and R. A. Grippi, Jr.	85 ✓
B. Nozzle Thrust Misalignment NASA Code 128-32-06-01, L. D. Strand	90
C. Prepolymer Functionality Determination Using a Model Polymerization System NASA Code 128-32-05-10, H. E. Marsh and J. J. Hutchison	95
D. Foams Produced From Carboxyl-Terminated Hydrocarbons NASA Code 128-32-05-10, S. Anderson, J. J. Hutchison, and H. E. Marsh, Jr.	99
E. Transition From Deflagration to Detonation in Granular Solid Propellant Beds NASA Code 128-32-06-01, O. K. Heiney	102
XI. Polymer Research	106
A. Ethylene Oxide-Freon 12 Decontamination Procedure: Reactions in the Decontamination Chamber and Effective Air-Flush Periods NASA Code 186-58-13-09, S. H. Kalfayan and R. H. Silver	106
B. Thermally Stable Urethane Elastomers NASA Code 186-68-13-03, E. F. Cuddihy and J. Moacanin	108
C. A Relationship Between Maximum Packing of Particles and Particle Size NASA Code 128-32-05-02, R. F. Fedors	109 ✓
XII. Research and Advanced Concepts	117
A. Pressure Distribution Along the Wall of an Axisymmetric Second-Throat Diffuser for Ambient Temperature Air Flow NASA Code 128-31-06-08, R. F. Cuffel, P. F. Massier, and L. H. Back	117
B. Suitability of a Hollow Cathode for a 20-cm-diam Ion Engine NASA Code 120-26-08-01, E. V. Pawlik and D. J. Fitzgerald	119
C. Liquid-Metal MHD Power Conversion NASA Code 120-27-06-03, D. G. Elliott, L. G. Hays, and D. J. Cerini	125
D. Efficiency of Thermionic Diodes at Reduced Power Output NASA Code 120-27-06-14, J. P. Davis	129
E. Clustered Ion Engine Systems Studies NASA Code 120-26-08-01, T. D. Masek	131 ✓
XIII. Liquid Propulsion	135
A. The Liquid-Phase Mixing of a Pair of Impinging Sheets NASA Code 731-12-02-11, R. W. Riebling	135

Contents (contd)

SPACE SCIENCES DIVISION

XIV. Space Instruments	141
A. Sterilizable, Ruggedized Imaging System	
NASA Code 125-24-01-03, L. R. Baker	141
B. Photo Sensor Evaluation	
NASA Code 125-24-01-03, K. J. Ando and L. R. Baker	142
C. Studies on the Photoconducting Layer in Slow-Scan Vidicons	
NASA Code 125-24-01-03, K. J. Ando and L. R. Baker	146
XV. Science Data Systems	149
A. Piece-wise Linear Approximation of a Mass Spectrometer Sweep Voltage	
NASA Code 186-68-03-04, W. Spaniol	149
XVI. Lunar and Planetary Sciences	155
A. Solar Wind Interaction With Solids	
NASA Code 185-42-12-01, H. C. Lord	155
B. 1967 Radar Observation of Mars	
NASA Code 185-41-24-01, R. L. Carpenter	157
C. Possibility of Permafrost Features on the Martian Surface	
NASA Code 185-37-20-12, F. A. Wade and J. N. deWys	160
XVII. Bioscience	163
A. Picric Acid Stability in Aqueous Sodium Hydroxide as Related to the Biosatellite Mission	
NASA Code 189-55-02-02, J. P. Hardy and J. H. Rho	163
XVIII. Fluid Physics	167
A. The Stability of Viscous Three-Dimensional Disturbances in the Laminar Compressible Boundary Layer. Part II	
NASA Code 129-01-08-02, L. M. Mack	167
XIX. Physics	171
A. Mechanism of the Reaction of Atomic Oxygen With Olefins	
NASA Code 129-02-01-04, W. B. DeMore	171
B. Dyadic Analysis of the World Models of Cosmology	
NASA Code 129-02-07-02, F. B. Estabrook, H. D. Wahlquist, and C. G. Behr	173
C. Unitary Representations of the Restricted Poincaré Group From a Unified Standpoint	
NASA Code 129-02-07-02, J. S. Zmuidzinas and K. L. Phillips	174
D. Testing Analytic Models Against Compressed Spectral Data	
NASA Code 129-02-04-01, E. L. Haines, R. H. Parker, and R. Gouw	175

Contents (contd)

TELECOMMUNICATIONS DIVISION

XX. Communications Systems Research	181
A. Sequential Decoding With Decision-Directed Phase Estimation	
NASA Code 125-21-01-02, J. A. Heller	181
B. Synchronization of PCM Channels by the Method of Word Stuffing	
NASA Code 125-21-02-03, S. Butman	187
C. Factoring Polynomials Over Finite Fields	
NASA Code 125-21-01-01, R. J. McEliece	190
D. On Automorphism Groups of Block Designs	
NASA Code 125-21-01-01, R. E. Block	194
E. Phase-Locking to Noisy Oscillators	
NASA Code 150-22-11-08, R. C. Tausworthe	198
F. Analysis of the Effect of Input Noise on a VCO	
NASA Code 150-22-11-08, R. M. Gray and R. C. Tausworthe	203
G. On S/N Estimation	
NASA Code 150-22-11-09, J. W. Layland	209
H. Digital Filtering of Random Sequences	
NASA Code 150-22-11-09, G. Jennings	213
I. The ϵ -Entropy of Certain Singular Measures on the Real Line	
NASA Code 150-22-17-08, T. S. Pitcher	221
XXI. Communications Elements Research	227
A. RF Techniques: 90-GHz Millimeter Wave Work	
NASA Code 125-21-03-04, W. V. T. Rusch, S. D. Slobin, and C. T. Stelzried	227
B. Quantum Electronics: Optical Communications Components	
NASA Code 125-22-02-01, M. S. Shumate and J. C. Siddoway	228
C. Low Noise Transponder Preamplifier Research	
NASA Code 150-22-17-01, S. M. Petty	232
D. Spacecraft Antenna Research	
NASA Code 186-68-04-02, R. M. Dickinson and K. Woo	233
E. RF Breakdown Studies: Multipacting Breakdown in Coaxial Transmission Lines 150–800 MHz	
NASA Code 125-22-01-02, R. Woo	240
XXII. Spacecraft Telemetry and Command	243
A. Multiple Mission Telemetry System: System Verification and Testing	
NASA Code 150-22-17-13, N. A. Burow and A. Vaisnys	243
B. Time Synchronization in an MFSK Receiver	
NASA Code 150-22-17-04, H. D. Chadwick	252

Contents (contd)

XXIII. Spacecraft Radio	265
A. High Impact S-Band Isolator Magnetic Materials Study	
NASA Code 150-22-17-06, A. W. Kermode	265
B. Effect of Interference on a Binary Communication Channel Using Known Signals	
NASA Code 186-68-04-11, M. A. Koerner	268
C. Spacecraft Power Amplifier Development Program	
NASA Code 186-68-04-09, L. J. Derr	278
D. Life Test Data Acquisition System	
NASA Code 186-68-04-09, R. S. Hughes	280
E. Low Data Rate Telemetry RF System Development	
NASA Code 150-22-17-06, R. B. Postal	284

ADVANCED STUDIES

XXIV. Future Projects	287
A. Lunar Ice	
J. R. Bruman and E. C. Auld	287
Abbreviations	291

I. Systems Analysis Research

SYSTEMS DIVISION

A. On the Statistics of Particle Fluxes Resulting From Solar Flares, P. Wesseling

The stream of high-energy particles emitted whenever a solar flare occurs impairs the efficiency of solar electric panels. With regard to solar panel design, it is important to know what the chances are that, after a given time, the total flux received will be below a given amount. The problem of deriving this probability is completely equivalent to a problem such as finding the probability that, after a given time, the damage done by lightning to a given city will be less than a given amount. Unfortunately, it turns out that a simple closed-form solution for this probability cannot be given.

It is assumed that the flares occur completely at random and are independent of each other. [This implies that the periodic (± 11 -yr) variations of the sun's activity are neglected; inclusion of these periodic variations would make the problem under consideration much more difficult.] From this assumption, it follows that the probability $p_{n,t}$ that n flares occur in a time t is given by a Poisson distribution:

$$p_{n,t} = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad (1)$$

The total flux Z received in a time t is given by

$$Z = X_1 + X_2 + \cdots + X_N \quad (2)$$

where N is distributed according to Eq. (1), and X_i is the flux generated by the i th flare in the sequence.

Let the probability density of the flare intensity be given by $q(x)$; i.e., the probability that the flux generated by the flare is between x and $x + dx$ is $q(x) dx$. Then, the probability density $P(z/n)$ of the total flux, given that n flares occurred, is given by

$$P(z/n) = \{q(z)\}^{n*} \quad (3)$$

where $\{q(z)\}^{n*}$ denotes the n -fold convolution of $q(x)$. This convolution is defined as follows:

$$\begin{aligned} \{q(z)\}^{n*} &= \{q(z)\} * \{q(z)\}^{(n-1)*} \\ &= \int_0^\infty q(x) \{q(z-x)\}^{(n-1)*} dx \end{aligned} \quad (4)$$

From Eqs. (3) and (1), it follows that the probability density of Z is given by

$$p_t(z) = \sum_{n=0}^{\infty} P(z/n) p_{n,t} = e^{-\lambda t} \sum_{n=0}^{\infty} \{q(z)\}^{n*} \frac{(\lambda t)^n}{n!} \quad (5)$$

The probability $P_t(m)$ that, after a time t , the total flux is less than m is given by

$$P_t(m) = \int_0^m p_t(z) dz = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} \int_0^m \{q(z)\}^{n*} dz \quad (6)$$

Numerical computations of $p_t(m)$ using Eq. (6) are hampered by the fact that $p_t(m)$ is given in the form of an infinite series, and even more by the need to compute the n -fold convolution of $q(x)$. The second complication can be eliminated, however, if the probability density of the flare intensity can be approximated by a gaussian density:

$$q(x) = \frac{1}{\sigma(2\pi)^{1/2}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} \quad (7)$$

The mean μ and the variance σ are to be determined such that Eq. (7) fits the observations in the best possible way.

Of course, there is no reason why, in reality, $q(x)$ should be gaussian. In fact, $q(x)$ cannot be gaussian, since, according to Eq. (7), there is a finite probability that some flares will generate negative fluxes. However, if μ is much larger than σ , this probability will be very small. It has to be decided on the basis of actual observations whether or not Eq. (7) is a good approximation.

If X_i is gaussian, Z is also gaussian; if it is known that n flares occurred, the mean of Z will be $n\mu$ and the variance $\sigma(n)^{1/2}$. Thus,

$$P(z/n) = \{q(z)\}^{n*} = \frac{1}{\sigma(2\pi n)^{1/2}} \exp \left\{ -\frac{(z - n\mu)^2}{2n\sigma^2} \right\} \quad (8)$$

Furthermore,

$$\int_0^m \{q(z)\}^{n*} dz \cong \frac{1}{2} \operatorname{erf} \left(\frac{m}{\sigma(2n)^{1/2}} - \frac{\mu(n)^{1/2}}{\sigma(2)^{1/2}} \right) + \frac{1}{2} \quad (9)$$

where

$$\operatorname{erf} x = \frac{2}{\pi^{1/2}} \int_0^x e^{-s^2} ds$$

and Eq. (8) has been integrated from $z = -\infty$ to $z = m$; hence,

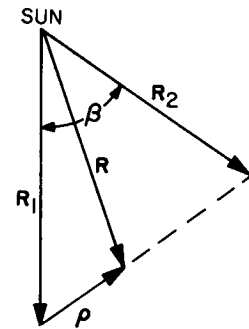
$$P_t(m) = \frac{1}{2} e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} \left\{ \operatorname{erf} \left(\frac{m}{\sigma(2n)^{1/2}} - \frac{\mu(n)^{1/2}}{\sigma(2)^{1/2}} \right) + 1 \right\} \quad (10)$$

B. An Alternate Expression for Light Time Using General Relativity, D. Holdridge

For the isotropic form of the Schwarzschild metric, the velocity of a radio signal is assumed to satisfy

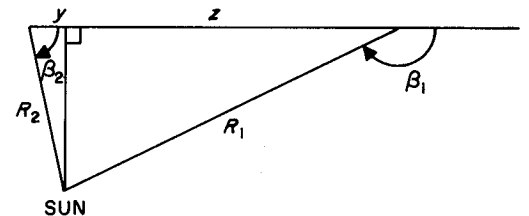
$$\dot{s} = c \left[1 - \frac{2\mu}{c^2 R} \right]$$

where R is the distance from the sun, μ is the gravitational constant of the sun, and c is the speed of light in vacuum. The signal is assumed to move in a straight line along $R_{12} = R_2 - R_1$, as shown below:



The expression for \dot{s} and the straight-line assumption neglect terms of order $1/c^4$.

With reference to the following geometry:



the light time from body 1 to body 2 is¹

$$T_2 - T_1 = \frac{R_{12}}{c} + \frac{2\mu}{c^3} \ln \left[\frac{(\tan \beta_1/2)}{(\tan \beta_2/2)} \right] \quad (1)$$

Another formula which has appeared in the literature (e.g., Ref. 1 and works by other authors such as I. I. Shapiro and D. O. Muhleman) has, as the logarithmic argument, the expression

$$\frac{[R_2 + y]}{[R_1 - z]} \quad (2)$$

Both of these formulas are, in a sense, somewhat artificial in that they involve angles and segments and only indirectly involve the quantities R_1 , R_2 , and R_{12} .

Presented here is an alternate formula which not only is simpler, but also involves in a symmetrical way the sides of the triangle:

$$T_2 - T_1 = \frac{R_{12}}{c} + \frac{2\mu}{c^3} \ln \left[\frac{(R_1 + R_2 + R_{12})}{(R_1 + R_2 - R_{12})} \right] \quad (3)$$

It is easy to show that Eq. (1) is equivalent to Eq. (3) by using the formulas for the tangents of the half-angles of a triangle:

$$\tan \frac{\beta_1}{2} = \cot \frac{(\pi - \beta_1)}{2} = \frac{(S - R_2)}{r}$$

$$\tan \frac{\beta_2}{2} = \frac{r}{(S - R_1)}$$

where

$$S = \frac{[R_1 + R_2 + R_{12}]}{2}$$

$$r^2 = \frac{[(S - R_1)(S - R_2)(S - R_{12})]}{S}$$

$$\frac{\tan \beta_1/2}{\tan \beta_2/2} = \frac{(S - R_1)(S - R_2)}{r^2}$$

$$= \frac{S}{(S - R_{12})}$$

$$= \frac{(R_1 + R_2 + R_{12})}{(R_1 + R_2 - R_{12})}$$

The logarithmic argument in expression (2) is also given by

$$\frac{(R_2 + y)}{(R_1 - z)} = \frac{(R_1 + z)}{(R_2 - y)}$$

since the roles of body 1 and body 2 may be interchanged by symmetry of the problem. To show that this is equivalent to the argument in Eq. (3), it is necessary merely to add numerators and denominators:

$$\frac{(R_2 + y)}{(R_1 - z)} = \frac{[R_1 + R_2 + (y + z)]}{[R_1 + R_2 - (y + z)]}$$

But $y + z = R_{12}$, so the argument is the same as that in Eq. (3).

To derive Eq. (3), we have, referring to the above sketch defining \mathbf{R}_1 and \mathbf{R}_2 :

$$\mathbf{R} = (1 - \alpha) \mathbf{R}_1 + \alpha \mathbf{R}_2, \quad 0 \leq \alpha \leq 1$$

$$R^2 = (1 - \alpha)^2 R_1^2 + \alpha^2 R_2^2 + 2\alpha(1 - \alpha) \mathbf{R}_1 \cdot \mathbf{R}_2$$

$$= \alpha^2 R_{12}^2 + 2\alpha(R_1 R_2 \cos \beta - R_1^2) + R_1^2$$

$$\rho = \alpha R_{12}$$

$$\dot{\rho} = \dot{\alpha} R_{12} = \frac{\rho}{\rho [c(1 - 2\mu/c^2 R)]}$$

$$\dot{\alpha} = \frac{c}{R_{12}} (1 - 2\mu/c^2 R)$$

Neglecting terms of order $1/c^4$,

$$\begin{aligned} T_2 - T_1 &= \frac{R_{12}}{c} \int_0^1 \left(1 + \frac{2\mu}{c^2 R} \right) d\alpha \\ &= R_{12}/c + 2\mu/c^3 \\ &\quad \times \ln \frac{R_2 + R_{12} + (R_1 R_2 \cos \beta - R_1^2)/R_{12}}{R_1 + (R_1 R_2 \cos \beta - R_1^2)/R_{12}} \end{aligned}$$

Making the substitution

$$R_1 R_2 \cos \beta - R_1^2 = \frac{[R_2^2 - R_1^2 - R_{12}^2]}{2}$$

¹Moyer, T. D., *Formulas for Link REGRES of the DPODP*, JPL Section 312 internal memorandum, Dec. 11, 1965.

the argument of the logarithm becomes

$$\frac{2R_2 R_{12} + R_{12}^2 + R_3^2 - R_1^2}{2R_1 R_{12} - R_{12}^2 + R_3^2 - R_1^2} =$$

$$\frac{(R_1 + R_2 + R_{12})(R_2 - R_1 + R_{12})}{(R_1 + R_2 - R_{12})(R_2 - R_1 + R_{12})}$$

The stated result then follows.

Reference

1. Ross, D. K., and Schiff, L. I., "Analysis of the Proposed Planetary Radar Reflection Experiment," *Phys. Rev.*, Vol. 141, No. 4, Jan. 1966.

C. A Simple Approach to Gravitational Theory,

H. Lass

1. Introduction

The flat space of special relativity theory must be abandoned when dealing with gravitational fields if Einstein's equivalence principle is accepted. Two observers, A and B, are considered at rest in flat space ($ds^2 = c^2 dt^2 - dx^2 - dy^2 - dz^2$) at a distance h apart. If the two observers are now given a uniform acceleration g , a doppler shift is recorded by B when A emits light signals to B. Within the order of $1/c^2$ terms, the change in frequency as given by $\Delta\nu/\nu = -gh/c^2$ can be calculated. From the equivalence principle, a corresponding doppler shift should be noted if A and B are at rest in a uniform gravitational field, namely $\Delta\nu/\nu = -\Delta\phi/c^2$, since $\Delta\phi = gh$. Thus, a clock at A should run at a different rate than a clock at B if A and B are at different gravitational potentials. This observation suggests a study of the line element $ds^2 = g_{\alpha\beta} dx^\alpha dx^\beta$, which led Einstein to his general theory of relativity. In Einstein's theory, the $g_{\alpha\beta}$ are determined from the field equations

$$R_\nu^\mu - \frac{1}{2} g_\nu^\mu R = -8\pi T_\nu^\mu \quad (1)$$

where T_ν^μ is the energy-momentum tensor, R_ν^μ is the Ricci tensor of Riemannian geometry, and $R = R_\mu^\mu$.

Of great advantage in Einstein's theory is the fact that the motions of particles, given by the geodesics associated

with the line element $ds^2 = g_{\alpha\beta} dx^\alpha dx^\beta$, namely,

$$\frac{d^2 x^i}{ds^2} + \Gamma_{jk}^i \frac{dx^j}{ds} \frac{dx^k}{ds} = 0 \quad (2)$$

are a consequence of the field equations.

A new approach to general relativity theory was proposed in Ref. 1. In this work, the $g_{\alpha\beta}$ are assumed to be functions of a scalar field ϕ , and the field equations corresponding to Eq. (1) are given by

$$R_\nu^\mu - \frac{1}{2} g_\nu^\mu R = -8\pi T_\nu^\mu \quad (3)$$

with

$$T_\nu^\mu = g_\nu^\mu L - \phi_{,\nu} \frac{\partial L}{\partial \phi_{,\mu}}$$

$$L = g^{\mu\nu} \phi_{,\mu} \phi_{,\nu}$$

It is then shown that the line element in isotropic coordinates is given by

$$ds^2 = e^{2\phi/c^2} dt^2 - e^{-2\phi/c^2} (dx^2 + dy^2 + dz^2) \quad (4)$$

for the static case, where ϕ satisfies the generalized Laplace's equation

$$e^{2\phi/c^2} \left(\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} \right) = 4\pi G \rho e^{2\phi/c^2} \quad (5)$$

with ρ the density of gravitating particles. Equation (5) must be posited in order that the geodesics of Eq. (2) be applicable.

For a central field (point source), $\phi = -GM/r$, it can be easily shown that the geodesics associated with the line element of Eq. (4) yield Einstein's value for the advance in the perihelion of the planet Mercury, and that the null geodesics yield Einstein's value for the bending of light. The red shift is a consequence of the form of ds^2 above.

Here it is assumed that the space-like term of ds^2 is conformal to a Euclidean 3-space, so that

$$ds^2 = c^2 F(\phi) dt^2 - G(\phi) (dx^2 + dy^2 + dz^2) \quad (6)$$

with F and G unknown functions of a scalar field ϕ . The functions F and G will be determined from the equivalence of inertial and gravitational masses for a particle

instantaneously at rest, and from the quantum-mechanical relationship $m_i c^2 = h\nu$. The field equation for ϕ will require an additional postulate. Finally, a simple cosmology will be presented.

2. Equations of Motion: Inertial Mass and Gravitational Mass

It is known that the geodesics corresponding to the line element of Eq. (6) can be obtained from the Lagrangian

$$L = \frac{m_0}{2} [G(\phi) (\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - c^2 F(\phi) \dot{t}^2] \quad (7)$$

with m_0 the scalar rest mass of a particle in the absence of a potential field; $\dot{x} = dx/ds$, etc.; and $\delta f L ds = 0$. Only the x -equation is considered:

$$\frac{d}{ds} \left(\frac{\partial L}{\partial \dot{x}} \right) = \frac{\partial L}{\partial x} \quad (8)$$

which yields

$$\frac{d}{ds} \left[m_0 G(\phi) \frac{dx}{ds} \right] = - \frac{m_0}{2} \frac{\partial \phi}{\partial x} [c^2 F'(\phi) \dot{t}^2 - (\dot{x}^2 + \dot{y}^2 + \dot{z}^2) G'(\phi)] \quad (9)$$

From ds^2 , the proper time for an observer at rest at (x, y, z) is given by $dT = F^{1/2} dt$, and proper distances for this observer are given by $dX = G^{1/2} dx$, $dY = G^{1/2} dy$, and $dZ = G^{1/2} dz$. Equation (9) expressed locally in terms of dT , dX , dY , dZ , and

$$v^2 = \left(\frac{dX}{dT} \right)^2 + \left(\frac{dY}{dT} \right)^2 + \left(\frac{dZ}{dT} \right)^2$$

takes the form

$$\frac{d}{dT} \left[\frac{m_0 G^{1/2}}{[1 - (v^2/c^2)]^{1/2}} \frac{dX}{dT} \right] = \frac{m_0 G^{1/2}}{2 [1 - (v^2/c^2)]^{1/2}} \left[\frac{c^2 F'(\phi)}{F(\phi)} - \frac{v^2 G'(\phi)}{G(\phi)} \right] \left(- \frac{\partial \phi}{\partial X} \right) \quad (10)$$

It must be understood that Eq. (10) represents only the local instantaneous motion of a particle as seen by an observer at rest at the point (x, y, z) . Interpreting this equation as stating that the local time rate of change of momentum is equal to the local gravitational force leads

to the following definitions of the inertial mass m_i and gravitational mass m_g :

$$m_i = \frac{m_0 [G(\phi)]^{1/2}}{[1 - (v^2/c^2)]^{1/2}} \quad (11)$$

$$m_g = \frac{m_0}{2} \frac{[G(\phi)]^{1/2}}{[1 - (v^2/c^2)]^{1/2}} \left[c^2 \frac{F'(\phi)}{F(\phi)} - v^2 \frac{G'(\phi)}{G(\phi)} \right]$$

so that instantaneously

$$\frac{d}{dT} \left(m_i \frac{dX}{dT} \right) = -m_g \frac{\partial \phi}{\partial X} = F_x \quad (12)$$

Under the assumption that $m_i = m_g$ for a particle instantaneously at rest ($v = 0$),

$$\frac{1}{2} c^2 \frac{F'(\phi)}{F(\phi)} = 1 \quad (13)$$

is obtained, the solution of which is $F(\phi) = e^{2\phi/c^2}$ with $F(0) = 1$.

From $m_i c^2 = h\nu$, it is noted that the frequency associated with an atomic clock ($v = 0$) is proportional to $[G(\phi)]^{1/2}$, so that proper time should be proportional to $[G(\phi)]^{-1/2}$. From $dT = F^{1/2} dt$, it is necessary to choose $[G(\phi)]^{-1} = F(\phi)$, with $G(0) = 1$. Hence, $G(\phi) = e^{-2\phi/c^2}$, yielding the line element of Eq. (4). From Eq. (11), it follows that

$$\frac{m_g}{m_i} = 1 + v^2/c^2 \quad (14)$$

Thus, if a particle is not at rest, its gravitational mass is larger than its inertial mass. It is interesting to note that, for photons ($v = c$), $m_g = 2m_i$. Application of this result to the bending of light, based on Newtonian theory, yields Einstein's value of 1.75'' of arc.

The field equation for the scalar ϕ is defined by

$$e^{-2\phi/c^2} \square \phi \equiv \frac{e^{-2\phi/c^2}}{|g|^{1/2}} \frac{\partial}{\partial x^\alpha} \left(|g|^{1/2} g^{\alpha\beta} \frac{\partial \phi}{\partial x^\beta} \right) = -4\pi G \rho_0 \quad (15)$$

so that

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} - \frac{1}{c^2} \frac{\partial}{\partial t} \left(e^{-4\phi/c^2} \frac{\partial \phi}{\partial t} \right) = 4\pi G \rho_0 \quad (16)$$

with ρ_0 the invariant scalar density of material as calculated by an observer moving with the material. For a

radial field, $\rho_0 = M\delta(x)\delta(y)\delta(z)$, so that $\phi = -GM/r$, which yields the Yilmaz line element (Ref. 1).

3. A Simple Cosmology

It is assumed that ρ_0 is given by $\rho_0(t) = \rho_{00}e^{\alpha t}$, so that the density of material in the universe increases with time (Bondi-Gold and Hoyle's creation of matter), and that there is no motion of matter. If ϕ is defined as $\phi = -c^2\beta t$, Eq. (16) yields $\alpha = 4\beta = 4(\pi G\rho_{00})^{1/2}$, with ρ_{00} the present density of material in the universe. Furthermore, $\dot{x} = \dot{y} = \dot{z} = 0$ satisfies the equations of the geodesics since $\phi = \phi(t)$. [The density of material created per unit time can be determined from $(\rho u^i)_{,i} = S/c$; this determination, however, is omitted here.]

The line element in spherical coordinates is

$$ds^2 = c^2 e^{-2\beta t} dt^2 - e^{2\beta t} (dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2) \quad (17)$$

The coordinate speed of light moving radially from $r = R$ to $r = 0$ (setting $ds = 0$, $d\theta = d\phi = 0$) is

$$\frac{dr}{dt} = -ce^{-2\beta t} \quad (18)$$

which yields

$$-R = \int_R^0 dr = -c \int_{t_T}^{t_R} e^{-2\beta t} dt \quad (19)$$

as an expression relating the coordinate time t_T at which light leaves a source at $r = R$ to the coordinate time t_R at which the light arrives at the origin $r = 0$.

A simple differentiation yields

$$e^{-2\beta t_R} \delta t_R = e^{-2\beta t_T} \delta t_T \quad (20)$$

relating the coordinate time interval δt_T between the departure of two wave crests from the source to the coordinate time interval δt_R between their arrival at the origin. Referring to proper time intervals yields

$$e^{-\beta t_R} \delta t_R^0 = e^{-\beta t_T} \delta t_T^0 \quad (21)$$

For convenience, $t_R = 0$ is chosen, so that

$$e^{-\beta t_T} \approx 1 + \frac{\beta R}{c}$$

from Eq. (19). Thus,

$$\begin{aligned} \delta t_R^0 &\approx \left(1 + \frac{\beta R}{c}\right) \delta t_T^0 \\ \frac{1}{v_R} &\approx \left(1 + \frac{\beta R}{c}\right) \frac{1}{v_T} \\ \frac{\delta v}{v} &= \frac{v_T - v_R}{v_T} \approx \frac{\beta R}{c} = \frac{R}{c\tau} \end{aligned} \quad (22)$$

with $\tau = 1/\beta$ defined as Hubble's constant.

Thus, to a first approximation, a doppler red shift occurs as proportional to the distance of particles from the origin, even though these particles remain at rest relative to the origin. The creation of matter yields a non-static line element, which in turn creates a doppler shift. (The coordinate speed of light changes continuously.) From $\beta = 1/\tau = (\pi G\rho_{00})^{1/2}$, $\rho_{00} = 1/\pi G\tau^2 \approx 3 \cdot 10^{-29}$ g/cm³ is obtained, an accepted value for the present density of material in the universe.

Referring to Eq. (22), a more accurate expression for $\Delta v/v$ can be shown to be

$$\begin{aligned} \frac{\Delta v}{v} &\approx \frac{R}{c\tau} e^{2\beta t_R} = \frac{R}{c\tau} e^{2t_R/\tau} \\ &\approx \frac{R}{c\tau^*} \end{aligned} \quad (23)$$

with

$$\tau^* \approx \frac{\tau}{\left(1 + \frac{2t_R}{\tau}\right)}$$

As t_R increases, Hubble's constant τ^* decreases.

Reference

1. Yilmaz, H., "New Approach to General Relativity," *Phys. Rev.*, Vol. III, No. 5, pp. 1417-1426, Sept. 1, 1958.

D. Introduction of New Planetary Masses Into Ephemeris Development Computations, J. D. Mulholland

The JPL Ephemeris Development Project is a continuing effort to improve the accuracy of the tabulated positions of the major planets and the Moon. In an earlier

article (SPS 37-45, Vol. IV, pp. 17-19), the author recommended a new set of planetary masses for use in JPL ephemeris development, based on new mass determinations and re-evaluations of past determinations. These masses are now being used in conjunction with simultaneous integrations of the major planets, with the epoch conditions differentially corrected to observational data. Since the differential correction process is based on linearity assumptions, it is desirable to take whatever action is possible to reduce the size of necessary corrections.

When the change from the old system of masses to the new set was introduced into the ephemeris computations, the available epoch conditions were based on an integration using the old masses. If they had been used without modification, this would have introduced an inconsistency which would have to be removed in the differential correction process. It arises in the following way: Considering each planetary orbit as a two-body motion, Kepler's third law requires that

$$a^3 n^2 = GS + GM_p \quad (1)$$

where a is the semi-major axis, n the mean motion, GS the solar gravitational constant, and GM_p the gravitational constant for the planet. In the perturbed planetary problem, where the Kepler ellipse no longer exists, this relationship may properly be regarded as *defining* the mean value of the semi-major axis. The mean motion is a directly observable quantity, obtained by a simple counting process, and is probably the most accurately known element of each of the planetary orbits. On the other hand, there is a direct linear relationship between the semi-major axis and the coordinates at any time. Thus, the adoption of a new value for M_p , used with epoch coordinates obtained with old masses, is tantamount to

changing the mean motion. If the mean motion is to be conserved while altering the masses, the coordinates and velocities at the epoch must also be altered by the multiplicative factor $1 + (\Delta a/a)$, where

$$\Delta a/a = \Delta M_p/3(1 + M_p)$$

and the solar mass is the mass unit. The masses and coordinate conversion factors that have been applied in current ephemeris development work are given in Table 1.

E. Collection of Optical and Radar-Range Data on the Planets, D. A. O'Handley

In the past few years, the various theories of motion for the major planets have been incorporated into the JPL development ephemerides. Two recently produced development ephemerides (DE 24 and DE 26) have been altered from the source theories through comparison of these theories with both optical and radar-range data and subsequent differential orbit correction.

The SSDPS, a series of programs which will numerically integrate the motion of the nine planets and correct this integration through evaluation of the residuals (observed - computed) is being checked out. To provide a set of observations which is as complete as possible over the time span 1949-1966 for a thorough checkout of the SSDPS, a systematic collection of planetary observations has been under way for a year. The general needs in setting up the standard format and the scope and type of data which will be used are discussed here.

The optical data for the period 1949-1966 have been obtained from various publications of the U.S. Naval Observatory (Refs. 1 and 2). In more recent times, the data are provided by punched cards in advance of publication.

There are approximately 5300 optical observations of the eight major planets (Pluto not included) for the period 1949-1966. These observations were recorded on punched cards in three different formats. The 6000 observations of Mars used in Clemence's work (Ref. 3) and the 600 observations of Pluto from 1930 to 1965 (Ref. 4) are in two additional formats. For the period 1913-1948, 17,000 observations have been added to the JPL files in 13 different card formats. It was necessary to adopt a format which contains as much information as possible from the various cards already on file and to make this format similar to a standard format for radar-range and doppler data. Two

Table 1. Coordinate conversion factors for new masses^a

Planet	$\Delta (M_p^{-1})$	M_p^{-1} (new values)	$(\Delta a/a) \times 10^{10}$
Mercury	17 000	6 017 000	-1.5696
Venus	504	408 504	-10.0798
Earth-Moon	-490	328 900	15.0764
Mars	5 000	3 098 500	-1.7388
Jupiter	0.0358	1 047.3908	-108.6787
Saturn	-2.4	3 499.2	652.7255
Uranus	61	22 930	-387.7380
Neptune	-243	19 071	2 198.9565
Pluto	0.0	400 000	0.0

^aThe epoch conditions are taken from Development Ephemeris 30 for Pluto and Development Ephemeris 26 for all other bodies.

formats² have been accepted by the U.S. Naval Observatory and the Naval Weapons Laboratory. This means that all transfer of planetary data will be in a single format henceforth.

The data set, which is currently being completed for use in the checkout of the SSDPS, has the optical data for the major planets taken from three series of meridian observations with the 6-in. transit circle of the U.S. Naval Observatory in the periods 1949-1955, 1956-1962, and 1963-1965. A current data series has been transmitted in punched cards to cover the period January 11, 1965 to April 11, 1966. Table 2 gives the number and distribution of the optical data for the planets. These data will soon be available in the new format.

Some of the radar data on cards at JPL have been used in the preparation of DE 24 and DE 26. The current attempt to collect all available radar data has resulted in acquisition of radar-range and doppler data for Mercury, Venus, and Mars from the Arecibo Ionospheric Observatory of Cornell-Sydney University Astronomy Center, the Millstone and Haystack sites of the Massachusetts Institute of Technology, and DSS 13 of the Deep Space Network.

The radar-range data used in checking out the SSDPS are given in Table 3. These data will be utilized in a single format also. The current data set is the beginning of a much more extensive collection of planetary data. The data given in Table 3 will be combined with data yet to be collected. The needs of JPL for past data will involve a step-by-step evaluation of each segment as it

²The formats and their codes are the subjects of JPL Section 311 internal memoranda.

Table 2. Number and distribution of optical data for the planets

Planet	Number of observations				
	1949-1955	1956-1962	1963-1965	1965-1966	Total
Sun	905	805	239	135	2084
Mercury	196	242	80	30	548
Venus	872	—	174	79	1125
Mars	72	81	51	30	234
Jupiter	99	142	62	29	332
Saturn	114	122	48	28	312
Uranus	99	127	65	36	327
Neptune	105	139	52	19	315

is added. The accuracy of radar may show that the optical data are not as necessary as is presently thought for obtaining certain elements of the orbit.

References

1. *Publications of the United States Naval Observatory* (Washington, D.C.), Series 2: Vol. XI, pp. 153-179, 1927; Vol. XIII; Vol. XVI, Pt. I, pp. 59-203, 1949; Vol. XVI, Pt. III, pp. 397-445, 1952; and Vol. XIX, Pt. I, pp. 49-110, 1964.
2. *Observations of the Sun, Moon, and Planets; Six-inch Transit Circle Results*, U.S. Naval Observatory Circulars: 103 (1956-62), Oct. 1964; 105 (1963-64), Nov. 1964; 108 (July 7, 1964 to Dec. 24, 1964), July 1965; and 115 (Jan. 11, 1965 to Apr. 11, 1966), Feb. 1967.
3. Clemence, G. M., *Astron. Papers*, Vol. 16, Pt. 2, pp. 261-333, 1961.
4. Cohen, C. J., Hubbard, E. C., and Oesterwinter, C., "New Orbit for Pluto and Analysis of Differential Corrections," *Astron. J.*, Vol. 72, No. 8, pp. 973-988, Oct. 1967.
5. Pettengill, G. H., Dyce, R. B., and Campbell, D. B., "Radar Measurements at 70 cm of Venus and Mercury," *Astron. J.*, Vol. 72, No. 3, pp. 334, 335, Apr. 1967.
6. Dyce, R. B., Pettengill, G. H., and Sanchez, A. D., "Radar Observations of Mars and Jupiter at 70 cm.," *Astron. J.*, Vol. 72, No. 6, p. 775, Aug. 1967.

Table 3. Radar-range data used to check out SSDPS

Planet	Period	Number of points	Source
Mercury	Apr. 7 to Sept. 13, 1964	17	Arecibo ^c
	Mar. 18 to Sept. 20, 1965	61	Arecibo ^c
	Mar. 9 to Aug. 25, 1966	46	Arecibo ^d
	May 26 to Aug. 8, 1967	27	Arecibo ^d
Venus	May 25 to July 31, 1964	1081	DSS 13 ^e
	Mar. 26 to Oct. 27, 1964	50	Arecibo ^c
	Dec. 15, 1965 to Feb. 22, 1966	98	DSS 13 ^e
	May 1 to Aug. 20, 1965	21	Arecibo ^c
	May 6 to Sept. 23, 1966	30	Arecibo ^d
	Mar. 16 to Oct. 19, 1967	79	Arecibo ^d
	Aug. 22 to Oct. 19, 1967	16	Millstone ^f
	July 12 to Sept. 14, 1967	14 ^a	Haystack ^f
Mars	July 12 to Sept. 14, 1967	15 ^b	Haystack ^f
	Nov. 19 to Dec. 17, 1964	6	Arecibo ^g
	Jan. 21 to June 3, 1965	33	Arecibo ^g

^a60-μs band data.
^b24-μs band data.
^cRef. 5.
^dPrivate communication from Arecibo Ionospheric Observatory, Oct. 1967.
^eLawson, C. L., and Holdridge, D. B., *Compression of Jet Propulsion Laboratory Venus Radar Data*, JPL Section 314 internal memorandum, Feb. 3, 1967.
^fPrivate communication from Massachusetts Institute of Technology, Lincoln Laboratory, Oct. 1967.
^gRef. 6.

F. On a New Necessary Condition for Optimal Control Problems With Discontinuities,

P. Dyer and S. R. McReynolds

1. Introduction

In SPS 37-46, Vol. IV, pp. 9-13, a new necessary condition is claimed for optimal control problems with discontinuities. Unfortunately, the dynamic programming derivation of the result contained errors, making the final result incorrect. The correct result is

$$\ddot{V} = \Delta f^T V_{xx}^+ \Delta f + H_x^+ \Delta f - \Delta H_x f - \Delta H_t \leq 0 \quad (1)$$

where

$$\Delta f = f^+ - f^-$$

$$\Delta H = H^+ - H^-$$

$$H = \frac{\partial V}{\partial x} f + L$$

This condition is to hold for maximization problems and is the same as Eq. (44) in SPS 37-46, except that the inequality sign was reversed. [In SPS 37-48, $\ddot{V} = -S$ is correctly defined by Eq. (34).] For minimization problems, the inequality sign in Eq. (1) should be reversed.

In this article, a correct dynamic programming derivation of this result and an analytic example to verify the correctness of the result are presented. This example appears to indicate that a similar result derived by Reid (Ref. 1) is in error.

2. Derivation of the Result

The same notation as in SPS 37-46 will be used: t^* will be used to denote a fixed time at which a jump discontinuity in the optimal control function occurs; $V^+(x, t)$ shall be used to denote the optimal return function on the right-hand side of the discontinuity, as well as its analytic extension obtained by using a continuous differentiable control law. The term $V^+(x, t)$, which satisfies the partial differential equation

$$V_t^+ + V_x^+ f^+ + L^+ = 0 \quad (2)$$

shall be required to have continuous third partials with respect to x and continuous second partials with respect to t ; f and L shall be assumed to be twice differentiable with respect to x and once differentiable with respect to t . These assumptions shall permit differentiation of

Eq. (2) in order to obtain relationships between the partial derivatives of $V^+(x, t)$ as needed.

Now, t_0 is defined to be a time prior to t^* , at which time the value of the state is chosen to be x_0 . The time of the discontinuity is allowed to vary; the variable time shall be denoted by $t_1 = t^* + dt$, where dt is small and may be positive or negative. The term $V^-(x_0, t_0, t_1)$ shall denote the value of the performance index for which the initial time is t_0 , the initial state is given by x_0 , and the discontinuity is at t_1 . The following relationship holds

$$V^-(x_0, t_0, t_1) = \int_{t_0}^{t_1} L(x, u^-, t) dt + V^+(x(t_1), t_1) \quad (3)$$

The above expression shall now be expanded around the nominal solution $t_1 = t^*$:

$$\begin{aligned} V^-(x_0, t_0, t_1) &= V^+(x(t^*), t^*) + \int_{t_0}^{t^*} L^- dt + (L^- + V_t^+) dt \\ &\quad + V_x^+ \Delta x + \frac{1}{2} [L_{tt}^- + L_{tx}^- f + V_{tt}^+] \Delta t^2 \\ &\quad + V_{tx}^+ \Delta t \Delta x + \frac{1}{2} V_{xx}^+ \Delta x^2 \dots \end{aligned} \quad (4)$$

where

$$\Delta x = x(t_1) - x(t^*)$$

The term Δx satisfies

$$\Delta x = f \Delta t + \frac{1}{2} (f_t + f_x f) \Delta t^2 + \dots \quad (5)$$

Employing Eq. (5) in Eq. (4) yields

$$\begin{aligned} V^-(x_0, t_0, t_1) &= V^+(x(t^*), t^*) \\ &\quad + \int_{t_0}^{t^*} L^- dt + (L^- + V_t^+ + V_x^+ f) \Delta t \\ &\quad + \frac{1}{2} [L_{tt}^- + L_{tx}^- f + V_{tt}^+ + 2V_{tx}^+ f + f V_{xx}^+ f \\ &\quad + V_x^+ f_t + V_x^+ f_x f] \Delta t^2 \dots \end{aligned} \quad (6)$$

To convert this to the required form, it is necessary to find expressions for V_t^+ , V_{tt}^+ , and V_{tx}^+ : V_t^+ is given by Eq. (1) as

$$V_t^+ = -V_x^+ f^+ - L^+ \quad (7)$$

The partial of this identity with respect to x yields

$$V_{tx}^+ = -V_{xx}^+ f^+ - V_x^+ f_x^+ - L_x^+ \quad (8)$$

The partial of Eq. (7) with respect to t yields

$$V_{tt}^+ = -V_{tx}^+ f^+ - V_x^+ f_t^+ - L_t^+ \quad (9)$$

Eq. (8) may be used to eliminate V_{tx}^+ ; hence,

$$V_{tt}^+ = (f^+)^T V_{xx} f^+ + (V_x^+ f_x^+ + L_x^+) f^+ - V_x^+ f_t^+ - L_t^+ \quad (10)$$

Now, using Eqs. (7), (8), and (10) to eliminate V_t^+ , V_{tx}^+ , and V_{tt}^+ from Eq. (6) yields

$$\begin{aligned} V^-(x_0, t_0, t_1) &= V^+(x(t^*), t^*) + \int_{t_0}^{t^*} L^- dt \\ &\quad + \dot{V} dt + \frac{1}{2} \ddot{V} dt^2 + \dots \end{aligned} \quad (11)$$

where

$$\dot{V} = L^- - L^+ + V_x^+(f^- - f^+) \quad (12)$$

and \ddot{V} is given by Eq. (1).

An optimal solution requires that $\dot{V} = 0$ and $\ddot{V} \leq 0$. The condition $\dot{V} = 0$ is equivalent to the well-known condition that the Hamiltonian must be continuous; $\ddot{V} \leq 0$ is the new condition given by Eq. (1).

3. Jump Conditions for the Partial Derivative

The jump condition for V_{xx} derived in SPS 37-46 will be re-derived here, not assuming $\dot{V} = 0$; a jump condition for V_x will also be derived.

By letting $t_0 \rightarrow t^*$, Eq. (11) can be written

$$V^-(x, t^*) = V^+(x, t^*) + \dot{V} dt + \frac{1}{2} \ddot{V} dt^2 + \dots \quad (13)$$

Expanding the above expression around $x(t^*)$ and using δx to denote $x - x(t^*)$, the following second-order expansion is obtained:

$$\begin{aligned} V^-(x, t^*) &= V^+(x(t^*), t^*) + V_x^+ \delta x + \dot{V} dt + \frac{1}{2} \delta x^T V_{xx}^+ \delta x \\ &\quad + \dot{V}_x \delta x dt + \frac{1}{2} \ddot{V} dt^2 + \dots \end{aligned} \quad (14)$$

where

$$\dot{V}_x = -\Delta H_x - V_{xx}^+ \Delta f \quad (15)$$

Choosing dt to maximize the expansion in Eq. (14) yields

$$dt = -\ddot{V}^{-1} [\dot{V} + \dot{V}_x \delta x] \quad (16)$$

Substituting this expression back into Eq. (14) gives

$$\begin{aligned} V^-(x, t^*) &= V^+(x(t^*), t^*) - \dot{V}^2 \ddot{V}^{-1} \\ &\quad + [V_x^+ - \dot{V} \ddot{V}^{-1} \dot{V}_x] \delta x \\ &\quad + \frac{1}{2} \delta x^T [V_{xx}^+ - \ddot{V}^T \ddot{V}^{-1} \dot{V}_x] \delta x \end{aligned} \quad (17)$$

Thus,

$$V_x^- = V_x^+ - \dot{V} \ddot{V}^{-1} \dot{V}_x \quad (18)$$

$$V_{xx}^- = V_{xx}^+ - \ddot{V}^T \ddot{V}^{-1} \dot{V}_x \quad (19)$$

which are the desired relationships. It should be noted that along an optimal trajectory $\dot{V} = 0$, and hence V_x is continuous.

4. Example

An example will be given to demonstrate that: (1) the above formulas for \dot{V} and \dot{V}_x are correct and, (2) Reid's condition must be incorrect. The correctness of the above formulas is verified by checking the formulas with an independent calculation. If the example presented satisfies all the classical conditions for optimality (including Reid's) except the condition derived here, and it is demonstrated that this is not a locally optimal solution, Reid's condition will be proven invalid.

Consider the system

$$\begin{aligned} \dot{x}_1 &= x_2 + u, & x_1(0) &= x_1^0 \\ \dot{x}_2 &= -u, & x_2(0) &= x_2^0 \end{aligned} \quad (20)$$

and the maximization of J , where

$$J = \frac{1}{2} x_1(T)^2 + \frac{1}{2} x_2(T)^2 \quad (21)$$

and where the control, u , is bounded, i.e.,

$$|u| \leq \pm \Delta$$

The Hamiltonian, H , is given by

$$H = (p_1 - p_2)u + p_1 x_2 \quad (22)$$

which is a maximum when

$$u = \Delta \operatorname{sgn}(p_1 - p_2)$$

with

$$\begin{aligned} p_1(t) &= x_1(T) \\ p_2(t) &= x_2(T) - x_1(T)(t - T) \end{aligned} \quad (23)$$

Since the switching function is linear in t , it is clear that no more than a single switch may exist on an optimal solution.

Because of the simplicity of the equations, it is possible to express the terminal state in terms of the initial state and the single switching time, t^* ; e.g.,

$$\begin{aligned} x_1(T) &= x_1^0 + x_2^0 T - (T - 2t^*)\Delta - t^{*2}\Delta + \left(\frac{T - 2t^*}{2}\right)^2 \Delta \\ x_2(T) &= x_2^0 + (T - 2t^*)\Delta \end{aligned} \quad (24)$$

with u^- taken as $+\Delta$. Hence, the expression for Eq. (21) becomes

$$\begin{aligned} J &= \frac{1}{2} \left\{ x_1^0 + x_2^0 T - (T - 2t^*)\Delta - t^{*2}\Delta \right. \\ &\quad \left. + \left(\frac{T - 2t^*}{2}\right)^2 \Delta \right\}^2 + \frac{1}{2} \left\{ x_2^0 + (T - 2t^*)\Delta \right\}^2 \end{aligned} \quad (25)$$

The derivative of J with respect to t^* is given by

$$\begin{aligned} J_{t^*} &= 2\Delta \left\{ x_1^0(1 - T) \right. \\ &\quad + x_2^0 \left[-1 + T - T^2 \right] - 2T\Delta + \frac{3T^2\Delta}{2} - \frac{T^3\Delta}{2} \\ &\quad + t^* \left[x_1^0 + x_2^0 T + 4\Delta - 5\Delta T^2 + \frac{5}{2}\Delta T^2 \right] \\ &\quad \left. + 3\Delta(1 - T)t^{*2} + t^{*3}\Delta \right\} \\ &= 0 \end{aligned} \quad (26)$$

The second derivative of J with respect to t^* is then

$$\begin{aligned} J_{t^*t^*} = \ddot{V} &= 2\Delta \left\{ x_1^0 + x_2^0 T + 4\Delta - 5\Delta T + \frac{5}{2}\Delta T^2 \right. \\ &\quad \left. + 6\Delta(1 - T)t^* + 3\Delta t^{*2} \right\} \end{aligned} \quad (27)$$

and $J_{t^*x^0}$ is

$$J_{t^*x^0} = 2\Delta(1 - T + t^*, -1 + T - T^2 + t^*T) \quad (28)$$

However, in the following analysis, $J_{t^*x(t^*)}$ is required (as opposed to $J_{t^*x^0}$). But,

$$J_{t^*x(t)} = C J_{t^*x(0)}$$

where

$$C = \begin{bmatrix} 1 & t^* \\ 0 & 1 \end{bmatrix}$$

Hence,

$$\begin{aligned} J_{t^*x(t)} &= 2\Delta [1 - T + t^*, -1 + T - t^* - (T - t^*)^2] \\ &= \dot{V}_x \end{aligned} \quad (29)$$

Equations (27) and (29) will now be used to verify the formulas for \ddot{V} and \dot{V}_x (Eqs. 1 and 15).

The Ricatti variable P^+ is given by

$$\dot{P} = -\frac{\partial f^T}{\partial x} P - P \frac{\partial f}{\partial x}$$

with

$$P(T) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

or

$$P^+(t) = \begin{bmatrix} 1 & T - t \\ T - t & 1 + (T - t)^2 \end{bmatrix}$$

Also,

$$f^- - f^+ = 2\Delta \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

and

$$H_x^+ - H_x^- = 0$$

$$H_t^+ - H_t^- = 0$$

Hence,

$$\begin{aligned} \dot{V}_x &= 2\Delta \begin{bmatrix} 1 & T - t^* \\ T - t^* & 1 + (T - t^*)^2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ &= 2\Delta [1 - T + t^*, -1 + T - t^* - (T - t^*)^2] \end{aligned}$$

which agrees with Eq. (29). Similarly,

$$\begin{aligned}\ddot{V} &= (f^- - f^+)^T P^+ (f^- - f^+) - H_x^T (f^- - f^+) \\ &= 4\Delta^2 [1, -1] \begin{bmatrix} 1 & T - t^* \\ T - t^* & 1 + (T - t^*)^2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ &\quad - 2\Delta \left\{ x_1^o + x_2^o T - (T - 2t^*)\Delta - t^{*2}\Delta \right. \\ &\quad \left. + \left(T - \frac{2t^*}{2} \right)^2 \Delta \right\}\end{aligned}$$

and

$$\begin{aligned}\ddot{V} &= 2\Delta \left[x_1^o + x_2^o T + 4\Delta - 5T\Delta + \frac{5}{2} \right. \\ &\quad \left. \times T^2\Delta + 6\Delta(1 - T)t^* + 3t^{*2}\Delta \right] \quad (30)\end{aligned}$$

which again checks.

Now, the particular case when $x_1(T) = 1$, $x_2(T) = 0$, $T = 2$, and $\Delta = -1$ is considered. Proceeding via Pontryagin's principle, the control u is chosen such that

$$u = \text{sgn}(p_1 - p_2)$$

A switch occurs when $p_1 = p_2$, i.e., when $t = t^*$. In this case, $t^* = 1$ and $x_1^o = x_2^o = 0$. This trajectory completely satisfies Pontryagin's principle, since the Hamiltonian and the costate variables are continuous and u has been chosen correctly. Furthermore, Reid's condition

$$\dot{x}^+ \dot{p}^- - \dot{x}^- \dot{p}^+ > 0$$

gives

$$[x_2^+ + u^+, -u^+] \begin{bmatrix} 0 \\ -1 \end{bmatrix} - [x_2^- + u^-, -u^-] \begin{bmatrix} 0 \\ -1 \end{bmatrix} > 0$$

As $u^+ = -u^-$, the equation becomes

$$[-u^- - u^-, u^- + u^+] \begin{bmatrix} 0 \\ -1 \end{bmatrix} = -2u^-$$

but $u^- = -1$. Hence, $\dot{x}^+ \dot{p}^- - \dot{x}^- \dot{p}^+ = 2$, and Reid's condition is satisfied.

The evaluation of the second derivative of J (Eq. 30) gives (here, $\Delta = -1$)

$$\ddot{V} = 2[+4 + 10 - 10 - 6 + 3] = +2$$

which is positive, indicating that the solution is *not* a maximum. This result may be confirmed by evaluation of the return function (Eq. 25).

For the nominal t^* , J is given by

$$J = \frac{1}{2}[-(2 - 2t^*) - t^{*2} + (2 - 2t^*)^2] + \frac{1}{2}[2 - 2t^*]^2$$

Substituting $t^* = 1$ gives $J = 0.5$. Now, considering a small perturbation in t^* , $\Delta t = \pm 0.05$ gives

$$\begin{aligned}J &= \frac{1}{2} \left[\pm 0.1 - \left(\frac{1.05}{0.95} \right)^2 + 0.005 \right]^2 + \frac{1}{2} [0.1]^2 \\ &= \frac{1}{2} \left[\frac{0.9975}{0.9975} \right] + \frac{1}{2} [0.1]^2 \\ &= 0.502603125\end{aligned}$$

i.e., $J(t^* \pm \Delta t) > J(t^*)$. Clearly, there is not a maximum. Hence, the new necessary condition is shown to hold, while Reid's condition is shown to be in error.

Reference

1. Reid, W. T., "Discontinuous Solutions in the Non-Parametric Problem of Mayer in the Calculus of Variations," *Am. J. Math.*, Vol. 57, pp. 69-93, 1935.

II. Systems Analysis

SYSTEMS DIVISION

A. Analysis of De-orbit Maneuver Execution

Errors, E. G. Piaggi

1. Introduction

Descent to a planet's surface may be accomplished with a capsule placed on a descent trajectory from an orbit about the planet. This article shows the effects of errors in the initial conditions of this descent trajectory on entry and landing parameters of interest. Orbit-determination uncertainties and de-orbit maneuver execution errors that give rise to these deviations in initial conditions are considered both together and separately. It will be shown that the orbit-determination uncertainties contribute very little to the deviation in entry and landing parameters. The errors in initial conditions are mapped to entry- and landing-parameter errors in a Monte Carlo analysis. Histograms of the distributions of some parameters are presented. The error analysis is carried out in parametric fashion both for the entry parameters and the initial-condition variations.

2. Description of the Nominal De-orbit Trajectory

The descent trajectory begins following a de-orbit maneuver by the descent capsule; hence, the orbit prescribes the locus of possible injection points for the descent

or transfer trajectory. The present analysis considers that the entire trajectory from the de-orbit maneuver to landing takes place in a vacuum. However, an atmospheric height of 243.8 km is assumed (although it does not influence the trajectory), and certain parameters are computed at this point in the trajectory.

The de-orbit trajectory is assumed to be elliptical and may be determined completely by specifying six independent parameters. There are many parameters to choose from, but previous work¹ indicated the choice of the set used in this analysis. One parameter may be immediately fixed (and, hence, will not be considered) if it is assumed that the transfer de-orbit trajectory is in the plane of the orbiter. The remaining five initial parameters used to define the trajectory, along with parameters of interest at entry and encounter, are defined in Table 1 and Fig. 1.

3. Description of the Random Trajectory

The five initial-condition parameters HA , HP , ΔV , T , and η are treated as random variables. Random numbers from a generator having a normal distribution with a standard deviation $\sigma = 1$ and a mean of zero are multiplied by the appropriate σ values of the respective variables.

¹Internal publications by R. R. Stephenson and J. O. Light.

Table 1. Nomenclature for initial, entry, and landing parameters

Initial parameters	
HA	height above planet's surface of orbit's apoapsis
HP	height above planet's surface of orbit's periapsis
TA ^a	true anomaly at which capsule applies additional velocity increment (de-orbit)
T ^a	time from periapsis passage to TA
ΔV	magnitude of the maneuver velocity increment added to initiate de-orbit trajectory
η	maneuver orientation angle (angle made by ΔV with local horizontal)
Entry parameters	
α_e	angle of attack: angle between inertial direction of de-orbit maneuver and direction of inertial velocity vector at entry
V_e	inertial velocity of capsule at entry
γ_e	entry angle: flight path angle (angle between capsule inertial velocity vector and local horizontal) at entry altitude (800,000 ft or 243.8 km)
T_e	entry time: time from de-orbit point to capsule entry on ballistic vacuum trajectory
PER _e	PER (see below) at time of capsule entry
B _{LOOK_e}	B _{LOOK} (see below) at time of capsule entry
Landing parameters	
PER	landing site location: angle measured (positive in direction of orbiter) from direction of periapsis of last (prior to capsule de-orbit) satellite orbit to line intersecting a vacuum trajectory with the Martian surface
B _{LOOK}	bus look angle: look angle (positive in direction of orbiter) to spacecraft bus from capsule (referenced to local vertical) at time of impact on a vacuum trajectory
C _{LOOK}	capsule look angle: look angle (positive in direction of orbiter) to capsule from orbiter at time of capsule impact, measured from line parallel to line of apsides and going through location of the spacecraft at time of impact to spacecraft-capsule line
ρ	range at impact between capsule and orbiting spacecraft on a vacuum trajectory
SCT	spacecraft time: time to overhead passage of spacecraft as seen from landed capsule, negative indicating that spacecraft bus has already passed overhead by amount of time shown

^aFor given HA and HP, TA is interchangeable with the time from periapsis passage; TA is used for the nominal de-orbit trajectory, while T is used for the random trajectory.

The initial conditions for the random trajectory are then obtained by adding these random perturbations to the nominal initial conditions. It should be noted that here, as was the case with the nominal trajectory, no out-of-the-plane (of-the-orbiter) component is assumed. The entry and landing parameters of the randomly perturbed trajectory are then obtained in the same way as for the nom-

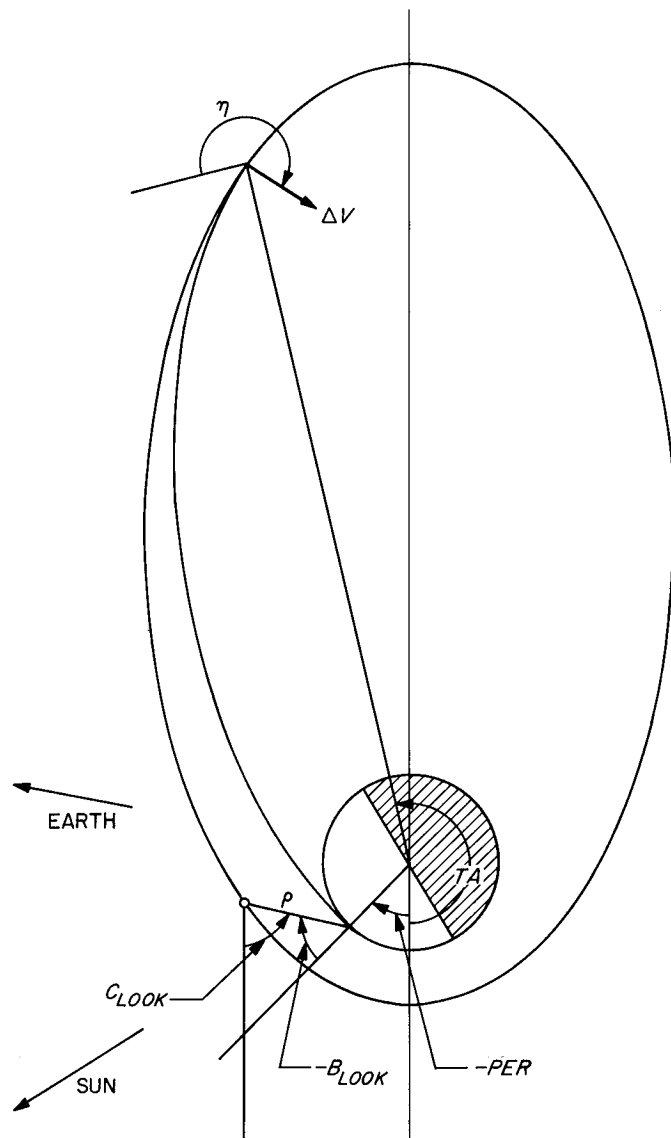


Fig. 1. Geometry of orbit and de-orbit trajectory

inal trajectory. By repeating this process a "large" number of times, the distribution of entry and landing parameters is obtained.

4. Numerical Results

This discussion is limited to numerical results for trajectories with $TA > 180$ deg. A future issue of the SPS, Vol. III, will give results for trajectories with $TA < 180$ deg and possibly a comparison of some salient points of the two types of trajectories.

The following analysis will first examine the effects of varying particular initial-condition parameters, as well as

certain entry and landing parameters. Secondly, it will establish which error sources seem to dominate the others in their effects on entry and landing parameters.

a. Variations of initial, entry, and landing parameters.

The descent trajectories analyzed here will, in general, be compared in pairs of trajectories which will differ markedly in only one of the following five parameters: HA , HP , ΔV , γ_e , and B_{LOOK} . It will be noted that only the first three of these are initial-condition parameters, while the last two are entry and landing parameters, respectively.

Of the five variables used to define the initial conditions of the de-orbit trajectories, the two that can be varied completely to suit particular entry and landing parameters are TA and η . Variations in the remaining three parameters are considered independently over their current likely range for a typical Mars mission, with the added constraints that γ_e be kept within approximately 14 to 20 deg and B_{LOOK} within approximately -30 to -60 deg.

Listed in Table 2 are trajectories for which HA ranges from 10,000 to 20,000 km, HP from 500 to 1500 km, and ΔV from 175 to 275 m/s; all of the trajectories given have γ_e and B_{LOOK} confined to the above limits.

Effects of variations in HA . The effects of a change in HA alone can be observed by comparing trajectories 1, 2, 3, and 4 with 5, 6, 7, and 8 (or trajectories 9, 10, 11, and 12 with 13, 14, 15, and 16). These two groups have a different HA (10,000 versus 20,000), but the same HP and ΔV . The four trajectories in each group are for both high and low γ_e values and more- and less-negative B_{LOOK} values. In comparing trajectories 5 and 1, which have high γ_e with more-negative B_{LOOK} values, it can be noted that the σ values of the entry and landing parameters examined are nearly equal. If the same comparison is carried out for the same conditions except for trajectories having low γ_e , i.e., trajectories 7 and 3, it can be noted that the trajectory with the high HA value has somewhat higher errors in most parameters examined, the largest differences being in the σ values for B_{LOOK} (1.7 versus 1.2 deg) and PER (2.59 versus 1.94 deg). It should also be noted that two parameters, namely, PER_e and γ_e , have a reversed behavior from the others; the difference here is small and may perhaps be due to the slightly lower γ_e of trajectory 3. By comparing the lower γ_e trajectories with the less-negative B_{LOOK} (for this group), it can be seen that the same behavior holds, with the differences in errors being larger than in any of the other three pairs of trajectories. The difference in σ for B_{LOOK} now amounts to 0.9 deg

(i.e., 2.3 deg for trajectory 8 and 1.4 deg for trajectory 4), while the difference in PER is 0.87 deg (i.e., 2.71 versus 1.84 deg). By comparing the corresponding four pairs of trajectories with a ΔV of 275 m/s, similar behavior may be observed.

It may be surmised from the above comparisons that a difference in HA from 10,000 to 20,000 km seems to have little effect for de-orbit trajectories with high γ_e . If a low γ_e is used, however, the 20,000-km HA has associated with it somewhat higher errors in most entry and landing parameters, especially B_{LOOK} and PER . The biggest difference in the errors of these parameters occurs for the less-negative nominal B_{LOOK} and low γ_e , the latter dominating the former.

Effects of variations in HP . The effects of significant differences in HP alone may be observed by comparing the σ values of the various entry and landing parameters of trajectories 13, 14, 15, and 16 with those of 17, 18, 19, and 20 (or 5, 6, 7, and 8 with 21, 22, 23, and 24). The general trend observed by comparing corresponding pairs of trajectories varying only in HP is that the σ values that seem most affected are those of α_e and V_e . The lower HP has associated with it the larger σ values of these two entry parameters. The σ values for some parameters such as ρ and SCT may differ somewhat, but not really significantly, especially as a percentage of the corresponding mean. It is of interest to note that the differences in α_e and V_e errors are somewhat more predominant for the trajectories with less-negative B_{LOOK} values. The above observations may be made whether trajectories 13, 14, 15, and 16 are compared with 17, 18, 19, and 20 or trajectories 5, 6, 7, and 8 with 21, 22, 23, and 24.

Effects of variations in ΔV . By comparing trajectories which differ only in ΔV (trajectories 1, 2, 3, and 4 with 13, 14, 15, and 16), it can be observed that errors in all except one parameter do not differ appreciably for comparable trajectories. The single exception is V_e . This exception can be well understood, since the errors in ΔV were taken to be proportional to the magnitude of ΔV , and these are seen to map directly into uncertainties in V_e . It can be noted that the trajectories with the less-negative B_{LOOK} values are associated with the greater differences (as high as a factor of 2.7 between trajectories 8 and 12, while only a factor of 1.7 between trajectories 7 and 11).

Effects of variations in B_{LOOK} . Trajectories which differ only in B_{LOOK} generally seem to have the same errors in entry and landing parameters, with the exception of ρ .

Table 2. Various de-orbit trajectories with approximately 14- to 20-deg γ_e and -30- to -60-deg B_{LOOK}

Trajectory	1	2	3	4	5	6	7	8
Initial parameters (nominal values) ^a								
HA, km	10,000	10,000	10,000	10,000	20,000	20,000	20,000	20,000
HP, km	500	500	500	500	500	500	500	500
ΔV , m/s	175	175	175	175	175	175	175	175
TA, deg	215	232.5	245	260	227.5	245	255	265
η , deg	224	184	230	188	231	193	227	184
Entry parameters (mean values) $\pm \sigma$								
γ_e , deg	19.92 ± 0.2	20.25 ± 0.18	13.96 ± 0.26	13.94 ± 0.28	19.94 ± 0.24	19.95 ± 0.23	14.22 ± 0.23	13.87 ± 0.23
B_{LOOK_e} , deg	-45.0 ± 0.4	-13.5 ± 0.4	-33.3 ± 0.5	-4.2 ± 0.5	-44.3 ± 0.4	-12.6 ± 0.4	-27.5 ± 0.5	-0.3 ± 0.5
PER_e , deg	-48.85 ± 0.47	-43.37 ± 0.47	-31.17 ± 0.67	-26.22 ± 0.77	-41.69 ± 0.51	-36.82 ± 0.55	-26.70 ± 0.53	-22.75 ± 0.55
T_e , h	1.3856 ± 0.0061	0.9698 ± 0.0065	0.7461 ± 0.0069	0.5667 ± 0.0065	1.3917 ± 0.0083	0.8645 ± 0.0084	0.6792 ± 0.0085	0.5576 ± 0.0081
α_e , deg	19.9 ± 0.6	71.6 ± 0.5	32.2 ± 0.7	84.3 ± 0.6	18.3 ± 0.6	68.9 ± 0.5	40.4 ± 0.7	89.6 ± 0.7
V_e , km/s	4.2970 ± 0.0007	4.2471 ± 0.0007	4.3041 ± 0.0010	4.2266 ± 0.0010	4.5298 ± 0.0007	4.4689 ± 0.0008	4.5165 ± 0.0010	4.4342 ± 0.0010
Landing parameters (mean values) $\pm \sigma$								
B_{LOOK} , deg	-59.7 ± 0.5	-30.0 ± 0.5	-59.8 ± 1.2	-30.2 ± 1.4	-59.9 ± 0.6	-29.8 ± 0.6	-57.4 ± 1.7	-29.8 ± 2.3
SCT, s	430 ± 5	115 ± 3	208 ± 7	71 ± 5	370 ± 4	97 ± 3	167 ± 8	65 ± 7
PER , deg	-36.34 ± 0.65	-31.22 ± 0.62	-7.72 ± 1.94	-3.80 ± 1.84	-28.84 ± 0.74	-24.08 ± 0.76	-0.44 ± 2.59	4.16 ± 2.71
ρ , km	1885 ± 14	903 ± 10	984 ± 19	583 ± 13	1741 ± 14	814 ± 10	857 ± 23	563 ± 19
C_{LOOK} , deg	84.0 ± 0.3	118.8 ± 0.5	112.5 ± 0.8	146.0 ± 0.8	91.3 ± 0.3	126.1 ± 0.5	122.2 ± 1.0	154.4 ± 0.8
Trajectory	9	10	11	12	13	14	15	16
Initial parameters (nominal values) ^a								
HA, km	20,000	20,000	20,000	20,000	10,000	10,000	10,000	10,000
HP, km	500	500	500	500	500	500	500	500
ΔV , m/s	275	275	275	275	275	275	275	275
TA, deg	240	262.5	267.5	285	227.5	257.5	257.5	282.5
η , deg	245	219	244	216	241	217	246	219
Entry parameters (mean values) $\pm \sigma$								
γ_e , deg	19.16 ± 0.29	19.98 ± 0.27	14.36 ± 0.22	14.26 ± 0.36	20.02 ± 0.24	19.73 ± 0.22	14.28 ± 0.28	14.06 ± 0.32
B_{LOOK_e} , deg	-43.4 ± 0.5	-13.5 ± 0.5	-28.0 ± 0.6	-3.6 ± 0.5	-44.1 ± 0.5	-13.4 ± 0.5	-33.1 ± 0.6	-5.4 ± 0.5
PER_e , deg	-39.54 ± 0.58	-34.75 ± 0.63	-26.21 ± 0.50	-21.21 ± 0.86	-48.44 ± 0.52	-38.93 ± 0.59	-31.21 ± 0.67	-23.65 ± 0.89
T_e , h	0.9297 ± 0.0084	0.5199 ± 0.0080	0.4794 ± 0.0083	0.3304 ± 0.0070	0.9932 ± 0.0061	0.5238 ± 0.0062	0.5429 ± 0.0066	0.3342 ± 0.0059
α_e , deg	15.4 ± 0.7	58.3 ± 0.7	35.3 ± 0.8	75.8 ± 0.9	14.9 ± 0.7	59.7 ± 0.6	28.4 ± 0.7	73.1 ± 0.8
V_e , km/s	4.5709 ± 0.0012	4.4866 ± 0.0018	4.5597 ± 0.0017	4.4409 ± 0.0027	4.3297 ± 0.0012	4.2563 ± 0.0016	4.3436 ± 0.0016	4.2301 ± 0.0023

Table 2 (contd)

Trajectory	9	10	11	12	13	14	15	16
Landing parameters (mean values) $\pm \sigma$								
B_{LOOK} , deg	-60.1 ± 0.7	-30.5 ± 0.6	-58.1 ± 1.6	-30.1 ± 2.1	-59.0 ± 0.6	-30.2 ± 0.6	-59.3 ± 1.2	-30.2 ± 1.3
SCT, s	337 ± 5	95 ± 3	170 ± 8	66 ± 7	414 ± 5	103 ± 3	207 ± 7	70 ± 5
PER, deg	-25.82 ± 0.90	-22.05 ± 0.88	0.00 ± 2.48	3.19 ± 3.12	-35.97 ± 0.73	-26.31 ± 0.78	-8.64 ± 1.87	-1.60 ± 2.06
ρ , km	1599 ± 18	783 ± 13	864 ± 23	568 ± 20	1830 ± 16	812 ± 11	985 ± 20	576 ± 14
C_{LOOK} , deg	94.0 ± 0.4	127.4 ± 0.8	121.9 ± 1.1	183.0 ± 1.3	85.0 ± 0.3	123.5 ± 0.7	112.1 ± 0.9	148.2 ± 1.1
Trajectory	17	18	19	20	21	22	23	24
Initial parameters (nominal values) ^a								
HA, km	10,000	10,000	10,000	10,000	20,000	20,000	20,000	20,000
HP, km	1500	1500	1500	1500	1500	1500	1500	1500
ΔV , m/s	275	275	275	275	175	175	175	175
TA, deg	212.5	205.0	237.5	208.5	210.0	201.0	217.5	195.0
η , deg	207	151	198	138	187	143	163	147.5
Entry parameters (mean values) $\pm \sigma$								
γ_e , deg	20.16 ± 0.19	20.13 ± 0.21	14.16 ± 0.24	13.85 ± 0.35	19.99 ± 0.18	15.94 ± 0.33	14.23 ± 0.25	19.97 ± 0.26
B_{LOOK_e} , deg	-47.0 ± 0.4	-14.3 ± 0.6	-38.5 ± 0.5	-0.5 ± 0.9	-46.3 ± 0.4	-6.0 ± 1.2	-33.6 ± 0.5	-13.0 ± 1.1
PER _e , deg	-44.49 ± 0.45	-40.37 ± 0.58	-20.88 ± 0.69	-22.41 ± 0.97	-38.37 ± 0.42	-29.15 ± 0.76	-22.53 ± 0.61	-39.88 ± 0.61
T_e , h	1.6839 ± 0.0051	2.1837 ± 0.0078	1.1697 ± 0.0056	2.2112 ± 0.0099	2.9633 ± 0.0078	4.1235 ± 0.0106	2.5649 ± 0.0089	4.7369 ± 0.0110
α_e , deg	29.8 ± 0.6	74.2 ± 0.3	46.2 ± 0.6	79.1 ± 0.3	41.4 ± 0.5	71.2 ± 0.3	62.8 ± 0.4	67.4 ± 0.3
Y_e , km/s	4.2859 ± 0.0008	4.2587 ± 0.0005	4.2601 ± 0.0010	4.2608 ± 0.0006	4.4991 ± 0.0003	4.4936 ± 0.0002	4.4824 ± 0.0003	4.4981 ± 0.0002
Landing parameters (mean values) $\pm \sigma$								
B_{LOOK} , deg	-60.1 ± 0.5	-30.1 ± 0.6	-60.5 ± 1.1	-30.5 ± 2.0	-60.1 ± 0.5	-30.2 ± 1.3	-60.0 ± 1.9	-24.7 ± 1.0
SCT, s	901 ± 10	301 ± 8	594 ± 15	239 ± 18	845 ± 10	239 ± 12	554 ± 22	293 ± 14
PER, deg	-32.21 ± 0.61	-28.10 ± 0.75	1.12 ± 1.57	0.94 ± 2.59	-25.62 ± 0.89	-10.49 ± 1.55	2.73 ± 2.55	-27.11 ± 0.86
ρ , km	3416 ± 20	2056 ± 21	2393 ± 32	1679 ± 26	3452 ± 24	1792 ± 27	2390 ± 45	2137 ± 38
C_{LOOK} , deg	87.7 ± 0.3	121.8 ± 0.8	120.7 ± 0.6	150.4 ± 1.1	94.3 ± 0.3	139.4 ± 1.4	122.7 ± 0.8	123.2 ± 1.4

^aThe σ values used for the initial parameters were as follows: $\sigma_{HA} = 10$ km, $\sigma_{HP} = 10$ km, $\sigma_{\Delta V} = 0.25\% \Delta V$, $\sigma_T = 10$ s, and $\sigma_\eta = 0.5$ deg.

It can be observed, however, that this is true only if the magnitudes of the errors are compared. Comparison shows that the errors as percentages of the nominal values are about the same, even for ρ . Therefore, trajectories that differ in B_{LOOK} only do not seem to have any discernible difference in the dispersions of the parameters. It can be noted, however, as pointed out above, different B_{LOOK} values may accentuate differences in other parameters if differences in some initial parameters are present.

Effects of variations in γ_e . By comparing trajectories with approximately the same HA , HP , ΔV , and B_{LOOK} but differing γ_e , it may be observed that the lower γ_e trajectories have parameters with the greater (or, at least, equal) dispersions. It seems from the data that, in general, the parameters whose σ values seem least affected by the variation in γ_e are the entry parameters. But even some of these, such as γ_e , B_{LOOK_e} , and PER_e , seem to give rise to higher σ values for lower γ_e and less-negative B_{LOOK} values. (As an example, trajectories 10 and 12 may be compared.)

The parameters that appear most affected by lower γ_e are B_{LOOK} , SCT , PER , and ρ . For example, trajectory 6 may be compared with trajectory 8; γ_e values for these trajectories are 19.95 and 13.87 deg, respectively. The corresponding errors in B_{LOOK} are 0.6 and 2.3 deg—nearly a factor of 4 difference. Similarly, the σ values for PER are 0.76 and 2.71 deg, respectively—again, nearly a factor of 4 difference.

It should be noted that variation in γ_e is the single most effective way of achieving PER variations. Hence, the PER values for trajectories 10 and 12 are quite different; i.e., the nominal values of PER are -22.05 and 3.19 deg, respectively. Since the less-negative PER values are achieved with lower γ_e , it may be concluded from the above that the less-negative PER values have, because of the lower γ_e values, greater errors in PER due to de-orbit maneuver execution errors.

b. Effects of error sources. The relative effects of the various error sources are directly dependent on the σ values used. Those that are currently considered to be somewhat realistic values for the 1970s for the five initial conditions used in this analysis are $\sigma_{HA} = 1$ km, $\sigma_{HP} = 0.3$ km, $\sigma_{\Delta V} = 0.25\%$ (ΔV), $\sigma_T = 5$ s, and $\sigma_\eta = 0.5$ deg.

To properly study the relative effects that the various error sources have on the entry and landing parameters, a Monte Carlo analysis was carried out, setting all but one

of the initial σ values equal to zero. Under these assumptions, the resulting dispersions of the parameters studied are shown in Table 3 along with the dispersions with all five nominal σ values. As can be readily seen, by far the dominating error source for all but one of the parameters is the pointing error in the de-orbit maneuver. (The one exception is T_e .) The error source that seems to be the next most important in contributing to entry- and landing-parameter errors is the proportional (or shut-off) error of the velocity. (This source of error is the chief contributor to dispersions in T_e .) It follows at once that, using the above σ values of the initial conditions, the orbit-determination errors (σ_{HA} , σ_{HP} , and part of σ_T) are negligible contributors to the dispersions of the entry and landing parameters examined.

The relative importance of error sources changes considerably, however, if the uncertainties in the initial conditions are those which were used to carry out the first part of this analysis. Table 4 shows the resulting dispersions when these values of the uncertainties (namely, $\sigma_{HA} = 10$ km, $\sigma_{HP} = 10$ km, $\sigma_{\Delta V} = 0.25\%$ (ΔV), $\sigma_T = 10$ s, and $\sigma_\eta = 0.5$ deg) are applied one at a time. It should be noted that here the dominating error source is the uncertainty in HP . The dominance here, however, is not as overwhelming as in the previous comparison; in fact, for some parameters the pointing error is still the largest contributor to the dispersions.

A complete units-of-variance analysis would be useful in determining the relative values of the uncertainties for which each uncertainty becomes the dominant error source. This analysis is not undertaken here, however.

c. Description of the dispersions. The dispersions in entry and landing parameters that arise from uncertainties in the initial conditions are more fully described if histograms of these dispersions can be constructed. The histograms of the distributions of some of the parameters investigated are shown in Figs. 2-4. As can be seen, these parameters (for this particular case) appear to be fairly normally distributed.

5. Conclusions

In summary, the following conclusions may be drawn from the preceding analysis:

- (1) The trajectory characteristic that most strongly influences landing-parameter errors is γ_e , with low γ_e giving rise to the largest errors.

Table 3. Relative effects of perturbations, using those σ values of initial parameters currently considered realistic for 1970s

Parameter	Value	σ_{HA} only	σ_{HP} only	$\sigma_{\Delta V}$ only	σ_T only	σ_η only
Initial	Nominal^a	Initial parameter σ values				
HA, km	18,000	1	0	0	0	0
HP, km	1000	0	0.3	0	0	0
ΔV , m/s	225	0	0	0.25%	0	0
TA, deg	230	0	0	0	5	0
η , deg	220	0	0	0	0	0.5
Entry	Mean $\pm \sigma$	Entry parameter σ values				
γ_e , deg	18.44 ± 0.16	0.005	0.01	0.06	0.01	0.15
B_{LOOK_e} , deg	-44.28 ± 0.41	0.41	0.003	0.07	0.01	0.39
PER_e , deg	-35.17 ± 0.28	0.01	0.01	0.12	0.03	0.24
T_e , h	1.4241 ± 0.0015	0.0005	0.0002	0.0007	0.0011	0.0005
α_e , deg	26.74 ± 0.61	0.01	0.01	0.06	0.01	0.59
V_e , km/s	4.4894 ± 0.0008	0.0000	0.0000	0.0000	0.0000	0.0008
Landing	Mean $\pm \sigma$	Landing parameter σ values				
B_{LOOK} , deg	-60.1 ± 0.5	0.003	0	0.1	0	0.5
SCT, s	518 ± 4	0.18	0	0	0	4
PER, deg	-20.78 ± 0.48	0.02	0.02	0.19	0.08	0.42
ρ , km	2299 ± 10	1	0	1	2	9
C_{LOOK} , deg	99.12 ± 0.11	0.02	0.01	0.06	0.04	0.08

^aThe σ values used for the initial parameters were as follows: $\sigma_{HA} = 1$ km, $\sigma_{HP} = 0.3$ km, $\sigma_{\Delta V} = 0.25\% \Delta V$, $\sigma_T = 5$ s, and $\sigma_\eta = 0.5$ deg.

Table 4. Relative effects of perturbations, using same σ values as those in Table 2

Parameter	Value	σ_{HA} only	σ_{HP} only	$\sigma_{\Delta V}$ only	σ_T only	σ_η only	$\sigma_\eta \sigma_{\Delta V}$ only
Initial	Nominal^a	Initial parameter σ values					
HA, km	18,000	10	0	0	0	0	0
HP, km	1000	0	10	0	0	0	0
ΔV , m/s	225	0	0	0.25%	0	0	0.25%
TA, deg	230	0	0	0	10	0	0
η , deg	220	0	0	0	0	0.5	0.5
Entry	Mean $\pm \sigma$	Entry parameter σ values					
γ_e , deg	18.44 ± 0.26	0.05	0.19	0.06	0.02	0.15	0.16
B_{LOOK_e} , deg	-44.3 ± 0.5	0.0	0.3	0.1	0.0	0.4	0.4
PER_e , deg	-35.17 ± 0.55	0.14	0.46	0.12	0.06	0.24	0.27
T_e , h	1.4242 ± 0.0074	0.0048	0.0053	0.0007	0.0022	0.0005	0.0009
α_e , deg	26.7 ± 0.6	0.1	0.2	0.1	0.0	0.6	0.6
V_e , km/s	4.4894 ± 0.0008	0.0002	0.0001	0.0000	0.0000	0.0008	0.0008
Landing	Mean $\pm \sigma$	Landing parameter σ values					
B_{LOOK} , deg	-60.1 ± 0.7	0.0	0.4	0.1	0.0	0.5	0.5
SCT, s	518 ± 7	2	5	0	1	4	4
PER, deg	-20.78 ± 0.87	0.21	0.69	0.19	0.09	0.42	0.46
ρ , km	2299 ± 18	7	14	1	3	9	9
C_{LOOK} , deg	99.1 ± 0.4	0.2	0.3	0.1	0.1	0.1	0.1

^aThe σ values used for the initial parameters were as follows: $\sigma_{HA} = 10$ km, $\sigma_{HP} = 10$ km, $\sigma_{\Delta V} = 0.25\% \Delta V$, $\sigma_T = 10$ s, and $\sigma_\eta = 0.5$ deg.

- (2) The maneuver execution errors (in particular, the pointing error) greatly dominate the contributions of the orbit-determination errors, using the current estimates of these error sources.

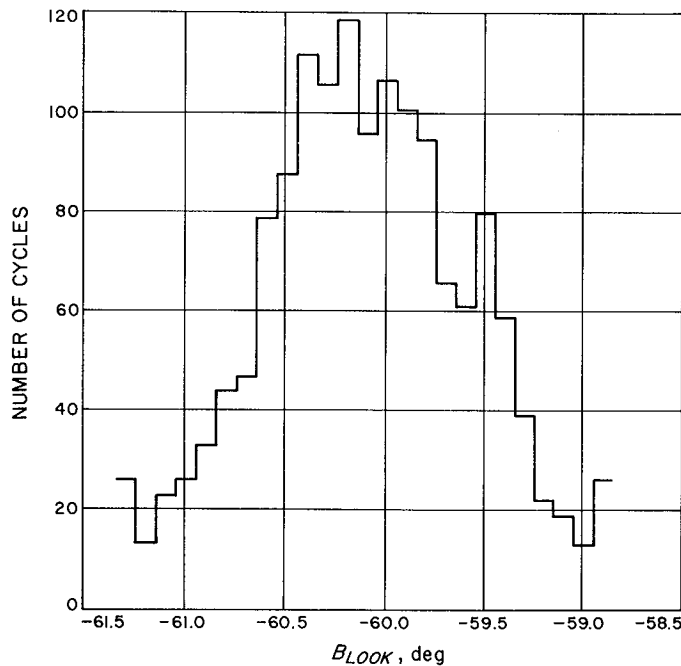


Fig. 2. Histogram of B_{LOOK} distribution

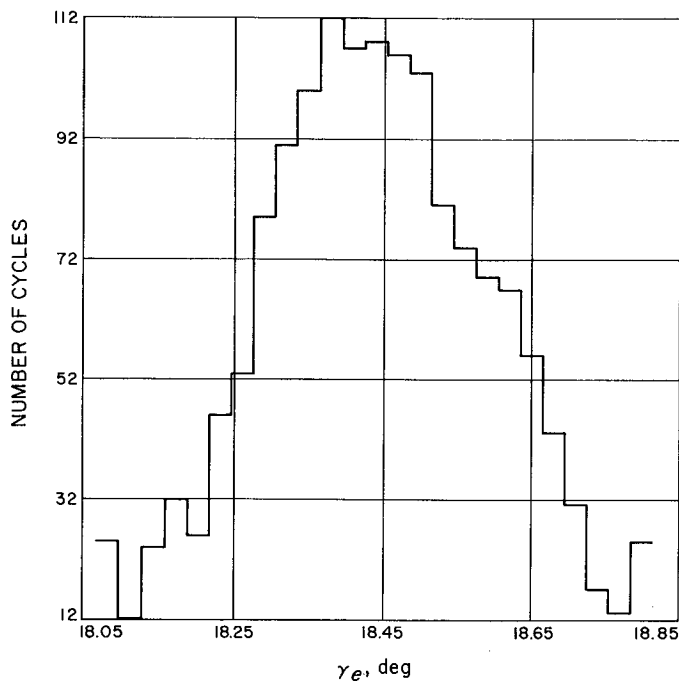


Fig. 3. Histogram of γ_e distribution

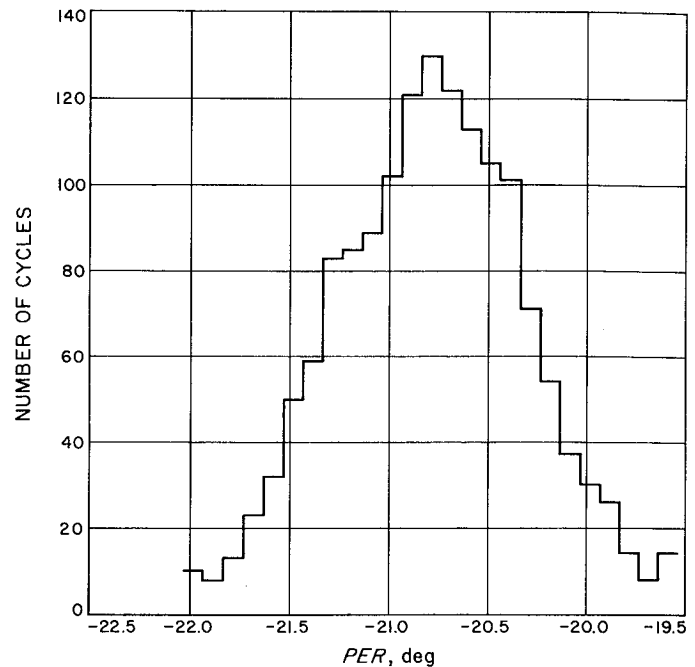


Fig. 4. Histogram of PER distribution

B. Analytical Determination of Midcourse Velocity Probability Distributions, L. Kingsland

1. Introduction

On interplanetary missions where large midcourse maneuvers may be required, the amount of propellant carried for midcourse maneuvers can comprise a significant part of the total mass of the spacecraft. In planning the amount of propellant to be carried on the spacecraft, an important consideration is the certainty that the propellant carried will be sufficient to perform all required midcourse maneuvers. In general, the probability function describing the expected distribution of midcourse velocity corrections is fairly difficult to determine.

The usual technique for determining the distribution of midcourse velocities is to run a Monte Carlo simulation of the maneuvers, based on a knowledge of the covariance matrices of each midcourse maneuver. Such a procedure is described, and several analytical distributions are fit to the data obtained from a Monte Carlo simulation in SPS 37-37, Vol. IV, pp. 1-11.

This article examines the mathematical structure of midcourse velocity distributions, describes how midcourse velocity distribution functions can be directly calculated by numerical integration, and discusses certain cases

which can be represented by means of an analytical distribution. The velocity correction maneuver following a hyperbolic encounter is examined in detail, and a simplified procedure for estimating the magnitude of such a maneuver is described.

2. Analysis

The expected distribution and orientation of a midcourse velocity maneuver is described, in general, by a covariance matrix, Λ_v , in a reference coordinate system. The correlation between components of the velocity in these reference coordinates can be eliminated by rotating by the matrix of eigenvectors, R :

$$\Lambda_u = R \Lambda_v R^T$$

where

$$\mathbf{u} = R\mathbf{v}$$

$$u = |\mathbf{u}| = |\mathbf{v}| = (u_x^2 + u_y^2 + u_z^2)^{1/2}$$

This rotation aligns the new coordinate system with the axes of symmetry of the dispersion ellipsoid, and, in these new coordinates, the components of \mathbf{u} will be uncorrelated. If the components of \mathbf{u} are normally distributed, the probability density function of each component will be:

$$f_i(u_i) = \frac{1}{\sigma_i (2\pi)^{1/2}} \exp \left[-\frac{1}{2} \left(\frac{u_i}{\sigma_i} \right)^2 \right], \quad i = x, y, z$$

Since the components of \mathbf{u} are uncorrelated, their three-dimensional probability density function (Ref. 1) will be the product of their individual probability density functions:

$$f(u_x, u_y, u_z) = \frac{1}{(2\pi)^{3/2} \sigma_x \sigma_y \sigma_z} \exp \left[-\frac{1}{2} \left(\frac{u_x^2}{\sigma_x^2} + \frac{u_y^2}{\sigma_y^2} + \frac{u_z^2}{\sigma_z^2} \right) \right]$$

By transforming into spherical coordinates (u, θ, ϕ)

$$u_x = u \cos \theta \cos \phi$$

$$u_y = u \sin \theta \cos \phi$$

$$u_z = u \sin \theta$$

and multiplying by the Jacobian of the transformation

$$J(u, \theta, \phi) = u^2 \cos \phi$$

the three-dimensional density function can then be expressed in terms of u , θ , and ϕ :

$$f(u, \theta, \phi) = \frac{u^2 \cos \phi}{(2\pi)^{3/2} \sigma_x \sigma_y \sigma_z} \exp \left[-\frac{u^2}{2} \left(\frac{\cos^2 \theta \cos^2 \phi}{\sigma_x^2} + \frac{\sin^2 \theta \cos^2 \phi}{\sigma_y^2} + \frac{\sin^2 \theta}{\sigma_z^2} \right) \right],$$

$$u \geq 0, 0 \leq \theta \leq \pi, -\frac{\pi}{2} \leq \phi \leq \frac{\pi}{2}$$

$$= 0 \text{ otherwise}$$

The density function of u is obtained by integrating over θ and ϕ :

$$f(u) = 8 \int_0^{\pi/2} \int_0^{\pi/2} f(u, \theta, \phi) d\theta d\phi, \quad u \geq 0$$

$$= 0, \quad u < 0 \quad (1)$$

The density function of the magnitude u of any midcourse velocity correction \mathbf{v} , whose covariance matrix is known and whose components are normally distributed, can be determined by numerical integration of Eq. (1). Numerical evaluation of this integral should be both simpler and more accurate than a Monte Carlo simulation, especially for probabilities very close to zero or unity. The distribution function, the mean, and the variance can be similarly calculated by numerical integration of the density function:

$$f(U) = \int_0^U f(u) du, \quad U \geq 0$$

$$= 0, \quad U < 0$$

$$\mu_u = E[u] = \int_0^\infty u f(u) du$$

$$E[u^2] = \int_0^\infty u^2 f(u) du$$

$$\sigma^2[u] = E[u^2] - E^2[u]$$

If the dispersion ellipsoid is spherical (i.e., if $\sigma_x = \sigma_y = \sigma_z \equiv \sigma_v$), the integral of Eq. (1) can be evaluated directly,

and the resulting density function will be Maxwell's distribution:

$$f(u) = \left(\frac{2}{\pi}\right)^{1/2} \frac{u^2}{\sigma_v^3} \exp\left[-\frac{1}{2}\left(\frac{u}{\sigma_v}\right)^2\right], \quad u \geq 0$$

$$= 0, \quad u < 0$$

The expectation of u^n will be:

$$E[u^n] = (2\sigma_v^2)^{n/2} \frac{2}{\pi^{1/2}} \Gamma\left(\frac{n+3}{2}\right)$$

where $\Gamma(\cdot)$ is the gamma function [$\Gamma(2) = 1$, $\Gamma(5/2) = 3\pi^{1/2}/4$, etc.]. Therefore,

$$\mu_u = E[u] = 2\sigma_v(2/\pi)^{1/2}$$

$$E[u^2] = 3\sigma_v^2$$

$$\sigma_u^2 = E[u^2] - E^2[u] = \left(3 - \frac{8}{\pi}\right)\sigma_v^2$$

The probability that the 3σ upper limit of Maxwell's distribution will be exceeded is:

$$\int_{\mu_u + 3\sigma_u}^{\infty} f(u) du = 0.00030$$

Another simplification of Eq. (1) will result if the velocity distribution is "pancake"-shaped (i.e., if two σ_i are equal and the third σ_i is zero). The resulting density function of u will then be Rayleigh's distribution:

$$f(u) = \frac{u}{\sigma_v^2} \exp(-u^2/2\sigma_v^2), \quad u \geq 0$$

$$= 0, \quad u < 0 \quad (2)$$

$$\mu_u = E[u^2] = \sigma_v \left(\frac{\pi}{2}\right)^{1/2}$$

$$E[u^2] = 2\sigma_v^2$$

$$\sigma_u^2 = E[u^2] - E^2[u]$$

$$= \left(2 - \frac{\pi}{2}\right)\sigma_v^2$$

The probability that the 3σ upper limit of Rayleigh's distribution will be exceeded is:

$$\int_{\mu_u + 3\sigma_u}^{\infty} f(u) du = 0.0055$$

It should be noted that the residual probability associated with the 3σ upper limit of the Rayleigh distribution is about 20 times larger than that of the Maxwell distribution. This comparison of the 3σ upper limits of two possible midcourse velocity distributions demonstrates that the term " 3σ " can be misleading if the reader attempts to associate a given level of certainty with that term. Whenever the term " 3σ " is used in connection with level of certainty, the associated probability or the distribution function should also be specified.

In certain cases, the components of a midcourse velocity vector may not necessarily be normally distributed. An example of such a situation is a midcourse maneuver which serves to correct for the execution errors of an earlier midcourse maneuver. In general, the velocity vector of such a maneuver is given by:

$$\mathbf{v}_2 = \left(\frac{\delta m}{\delta \mathbf{v}_2}\right)^{-1} \left(\frac{\delta m}{\delta \mathbf{v}_1}\right) \delta \mathbf{v}_1$$

where $\delta \mathbf{v}_1 = u_1 \mathbf{e}$, u_1 is the magnitude of the first midcourse maneuver, and \mathbf{e} is a normally distributed random vector. Since u_1 is a scalar,

$$\mathbf{v}_2 = u_1 \left(\frac{\delta m}{\delta \mathbf{v}_2}\right)^{-1} \left(\frac{\delta m}{\delta \mathbf{v}_1}\right) \mathbf{e} = u_1 \mathbf{w}$$

Each component of \mathbf{v}_2 will therefore be a product of u_1 and a linear function of the components of \mathbf{e} :

$$v_{2i} = u_1 (a_{1i} e_1 + a_{2i} e_2 + a_{3i} e_3)$$

$$= u_1 w_i$$

Since each component of \mathbf{e} is normally distributed, and since each component of \mathbf{w} is a linear function of the components of \mathbf{e} , it can be shown (Ref. 1) that each component of \mathbf{w} will be normally distributed:

$$f(w_i) = \frac{1}{(2\pi)^{1/2} \sigma_i} \exp\left[-\frac{1}{2}\left(\frac{w_i}{\sigma_i}\right)^2\right]$$

The probability density function of the product of two random, independent variables can be obtained by integrating the density functions of those two variables as

follows (Ref. 1):

$$\begin{aligned}
 f(v_{2i}) &= \int_{-\infty}^{\infty} \frac{1}{|u_1|} f_{u_1}(u_1) f_{w_i}\left(\frac{v_{2i}}{u_1}\right) du_1 \\
 &= \frac{8}{(2\pi)^{1/2} \sigma_i} \int_0^{\infty} \exp\left[-\frac{1}{2}\left(\frac{v_{2i}}{u_1 \sigma_i}\right)^2\right] \frac{du_1}{|u_1|} \int_0^{\pi/2} \int_0^{\pi/2} f(u_1, \theta, \phi) d\theta d\phi
 \end{aligned} \quad (3)$$

In general, Eq. (3) will require numerical integration, but, in the event that $f_{u_1}(u_1)$ is a Rayleigh density function, integration can be performed analytically:

$$\begin{aligned}
 f(v_{2i}) &= \frac{1}{(2\pi)^{1/2} \sigma_v^2 \sigma_i} \int_0^{\infty} u_1 \exp\left\{-\frac{1}{2}\left[\left(\frac{u_1}{\sigma_v}\right)^2 + \left(\frac{v_{2i}}{u_1 \sigma_i}\right)^2\right]\right\} \frac{du_1}{|u_1|} \\
 &= \frac{1}{2\sigma_v \sigma_i} \exp\left(-\frac{|v_{2i}|}{\sigma_i \sigma_v}\right)
 \end{aligned}$$

Assuming that the coordinate system has been rotated so that the components of \mathbf{v}_2 are independent, the joint probability density function of the components will be:

$$f(v_{2x}, v_{2y}, v_{2z}) = \frac{1}{8\sigma_v^3 \sigma_x \sigma_y \sigma_z} \exp\left\{-\frac{1}{\sigma_v} \left[\frac{|v_{2x}|}{\sigma_x} + \frac{|v_{2y}|}{\sigma_y} + \frac{|v_{2z}|}{\sigma_z}\right]\right\}$$

Transforming to spherical coordinates u_2 , θ , and ϕ as before:

$$f(u_2, \theta, \phi) = \frac{u_2^2 \cos \phi}{8\sigma_v^3 \sigma_x \sigma_y \sigma_z} \exp\left\{-\frac{u_2}{\sigma_v} \left[\frac{|\cos \theta \cos \phi|}{\sigma_x} + \frac{|\sin \theta \cos \phi|}{\sigma_y} + \frac{|\sin \phi|}{\sigma_z}\right]\right\}$$

The probability density function of u_2 , the magnitude of \mathbf{v}_2 , can then be obtained by numerical integration:

$$\begin{aligned}
 f(u_2) &= \frac{1}{\sigma_v^3 \sigma_x \sigma_y \sigma_z} \int_0^{\pi/2} \int_0^{\pi/2} \exp\left\{-\frac{u_2}{\sigma_v} \left[\frac{\cos \theta \cos \phi}{\sigma_x} + \frac{\sin \theta \cos \phi}{\sigma_y} + \frac{\sin \phi}{\sigma_z}\right]\right\} u_2^2 \cos \phi d\phi d\theta, \quad u_2 \geq 0 \\
 &= 0, \quad u_2 < 0
 \end{aligned}$$

3. Velocity Correction Following a Hyperbolic Encounter

In multiple-planet swingby missions, the largest velocity correction maneuvers are generally those required following a planetary hyperbolic encounter. For example, it has been shown (Ref. 2) that, for a 1970 mission to Mercury via a swingby at Venus, the post-Venus maneuver would account for about 80% of the total mission mid-course propellant requirements.

The deviation in the outbound relative velocity vector, $\mathbf{V}_{\infty o}$, resulting from variations in the inbound relative

velocity vector, $\mathbf{V}_{\infty i}$, and the vector aim point, \mathbf{B} , during a hyperbolic planetary encounter is given in Ref. 3 as:

$$\delta \mathbf{V}_{\infty o} = K \delta \mathbf{V}_{\infty i} + L \delta \mathbf{B}$$

where K and L are linear operators. It has been shown (Ref. 2) that target errors resulting from expected errors in \mathbf{B} will be significantly larger than those from expected errors in $\mathbf{V}_{\infty i}$:

$$\delta \mathbf{V}_{\infty o} \cong L \delta \mathbf{B} \quad (4)$$

where

$$L = -V_{\infty} \frac{\sin 2\nu}{b} I + \frac{\sin^2 2\nu}{b^2} V_{\infty i} \mathbf{B}^T + V_{\infty} \frac{\sin 2\nu}{b^3} \\ \times (1 + \cos 2\nu) \mathbf{B} \mathbf{B}^T$$

with ν as the total bending angle of the hyperbolic swingby. Expressed in RST coordinates:

$$\mathbf{B} = b \cos \theta \mathbf{i}_T + b \sin \theta \mathbf{i}_R$$

$$\delta \mathbf{B} = (\cos \theta \delta b - b \sin \theta \delta \theta) \mathbf{i}_T + (\sin \theta \delta b + b \cos \theta \delta \theta) \mathbf{i}_R$$

$$\mathbf{B}^T \delta \mathbf{B} = b \cos^2 \theta \delta b - b^2 \sin \theta \cos \theta \delta \theta + b \sin^2 \theta \delta b \\ + b^2 \sin \theta \cos \theta \delta \theta \\ = b \delta b$$

$$V_{\infty i} = V_{\infty} \mathbf{i}_S$$

Substituting the above relationships into Eq. (4):

$$\delta \mathbf{V}_{\infty o} = -V_{\infty} \frac{\sin 2\nu}{b} [(\cos \theta \delta b - b \sin \theta \delta \theta) \mathbf{i}_T \\ + (\sin \theta \delta b + b \cos \theta \delta \theta) \mathbf{i}_R] \\ + \frac{\sin^2 2\nu}{b^2} b \delta b V_{\infty} \mathbf{i}_S + \frac{V_{\infty} \sin 2\nu}{b^3} b \delta b (1 + \cos 2\nu) \\ \times (b \cos \theta \mathbf{i}_T + b \sin \theta \mathbf{i}_R) \\ = \frac{V_{\infty} \sin 2\nu}{b} [\sin 2\nu \delta b \mathbf{i}_S + (\cos \theta \cos 2\nu \delta b \\ + b \sin \theta \delta \theta) \mathbf{i}_T + (\sin \theta \cos 2\nu \delta b - b \cos \theta \delta \theta) \mathbf{i}_R]$$

It should be noted that $\delta \mathbf{V}_{\infty o}$ can be divided into two components, both perpendicular to the outgoing asymptote. One is in the trajectory plane and proportional to δb :

$$\frac{V_{\infty} \sin 2\nu}{b} \delta b [\sin 2\nu \mathbf{i}_S + \cos 2\nu (\cos \theta \mathbf{i}_T + \sin \theta \mathbf{i}_R)]$$

The other is perpendicular to the trajectory plane and proportional to $\delta \theta$:

$$V_{\infty} \sin 2\nu \delta \theta [\sin \theta \mathbf{i}_T - \cos \theta \mathbf{i}_R]$$

Therefore, the pre-encounter variations in the aim point vector map into a "pancake"-shaped ellipsoid perpendicular to the outgoing asymptote. This "pancake" effect

has been noted previously,² but the significant point here is that the required post-encounter velocity corrections will generally be perpendicular to the outgoing asymptote. This may be an important consideration in design problems related to the rotation of a spacecraft for propulsive maneuvers.

The covariance matrix of the outgoing relative velocity is:

$$\Lambda_{V_{\infty o}} = \overline{(\delta \mathbf{V}_{\infty o})(\delta \mathbf{V}_{\infty o})^T}$$

The three terms on the trace of the covariance matrix are:

$$\overline{(\delta \mathbf{V}_{\infty S})^2} = V_{\infty}^2 \frac{\sin^4 2\nu}{b^2} \overline{(\delta b)^2} = V_{\infty}^2 \frac{\sin^4 2\nu}{b^2} \sigma_b^2$$

$$\overline{(\delta \mathbf{V}_{\infty T})^2} = V_{\infty}^2 \frac{\sin^2 2\nu}{b^2} [\cos^2 \theta \cos^2 2\nu \overline{(\delta b)^2} + b^2 \sin^2 \theta \overline{(\delta \theta)^2} \\ + 2b \sin \theta \cos \theta \cos 2\nu \overline{\delta b \delta \theta}] \\ = \frac{V_{\infty}^2 \sin^2 2\nu}{b^2} [\cos^2 \theta \cos^2 2\nu \sigma_b^2 + b^2 \sin^2 \theta \sigma_{\theta}^2 \\ + 2b \sin \theta \cos \theta \cos 2\nu \overline{\delta b \delta \theta}]$$

$$\overline{(\delta \mathbf{V}_{\infty R})^2} = V_{\infty}^2 \frac{\sin^2 2\nu}{b^2} [\sin^2 \theta \cos^2 2\nu \sigma_b^2 + b^2 \cos^2 \theta \sigma_{\theta}^2 \\ - 2b \sin \theta \cos \theta \cos 2\nu \overline{\delta b \delta \theta}]$$

The sum of the trace of $\Lambda_{V_{\infty o}}$ is:

$$\frac{V_{\infty}^2 \sin^2 2\nu}{b^2} (\sigma_b^2 + b^2 \sigma_{\theta}^2)$$

but the bending angle ν is a function of V_{∞} and b :

$$\tan \nu = \frac{\mu}{b V_{\infty}^2}$$

Using trigonometric identities, it can be shown that

$$\sin 2\nu = \frac{2\mu}{b V_{\infty}^2} \left(1 + \frac{\mu^2}{b^2 V_{\infty}^4} \right)^{-1}$$

The sum of the trace can therefore be rewritten as follows:

$$\frac{V_{\infty}^2 \sin^2 2\nu}{b^2} (\sigma_b^2 + b^2 \sigma_{\theta}^2) = \frac{4\mu^2 (\sigma_b^2 + b^2 \sigma_{\theta}^2)}{b^4 V_{\infty}^2} \left(1 + \frac{\mu^2}{b^2 V_{\infty}^4} \right)^{-2}$$

²Long, J., et al., "Study of a 1973 Venus-Mercury Mission With a Venus Entry Probe," June 15, 1967 (JPL internal document).

The rms correction maneuver velocity is defined as the square root of the sum of the trace of $\Delta V_{\infty o}$:

$$(\Delta V \text{ rms}) = \frac{2\mu(\sigma_b^2 + b^2\sigma_\theta^2)^{1/2}}{b^2V_\infty} \left(1 + \frac{\mu^2}{b^2V_\infty^4}\right)^{-1} \quad (5)$$

Equation (5) can be checked against the numerical results obtained by Sturms and Cutting (Ref. 2) for a mission to Mercury via a close encounter with Venus, with launch on August 14, 1970, with the following hyperbolic encounter conditions:

$$V_\infty = 7.748 \text{ m/s}$$

$$b = 13,200 \text{ km}$$

$$\sigma_b = b\sigma_\theta = 100 \text{ km}$$

The rms velocity calculated from Eq. (5) is 57.6 m/s, which is identical to the conic analysis result obtained by Sturms and Cutting.

The relationship derived above provides a simple means of estimating the required velocity correction maneuver following a planetary hyperbolic encounter. It also provides a useful insight into the effect produced by V_∞ and b on post-encounter velocity deviations.

In the example cited above, the orbit-determination error is circular with $\sigma = 100 \text{ km}$. As a result of the foregoing analysis, it can be shown that the distribution of the post-Venus velocity correction will be two-dimensional normal and circular, with $\sigma = 57.6 \text{ m/s}$. Since the distribution of ΔV is two-dimensional normal, the magnitude of ΔV will obey Rayleigh's distribution (Eq. 2). This confirms the findings presented in SPS 37-37, which were derived by fitting analytical distributions to Monte Carlo simulations of the post-Venus midcourse maneuver.

4. Conclusions

The evaluation of midcourse velocity probability distributions by numerical integration appears feasible. In certain instances, especially in the case of probabilities near zero, it is possible that the use of numerical integration may be faster and more accurate than a Monte Carlo analysis.

References

1. Parzen, E., *Modern Probability Theory and Its Applications*. John Wiley & Sons, Inc., New York, 1960.
2. Sturms, F. M., and Cutting, E., "Trajectory Analysis of a 1970 Mission to Mercury Via a Close Encounter With Venus," Paper 65-90, presented at the AIAA 2nd Aerospace Sciences Meeting, New York, Jan. 25-27, 1965.
3. Battin, R. H., *Astronautical Guidance*. McGraw-Hill Book Co., Inc., New York, 1964.

III. Computation and Analysis

SYSTEMS DIVISION

A. Abstracts of Certain Mathematical Subroutines, I., R. Hanson

In September 1966, JPL began the collection and development of general-purpose mathematical computer subroutines. The areas selected for the initial phase of this project were matrix inversion, solutions of linear systems of algebraic equations, matrix pseudoinversion, eigenvalue-eigenvector calculations, real or complex roots of polynomial equations, "reliable" methods for nonlinear least squares problems and zeroes of vector valued functions of several variables, and interval arithmetic subroutines.

In this article we will present the abstracts of several FORTRAN IV callable subroutines which affect calculations in each of the above categories. We will also include the abstracts of subroutines which are nontrivial applications of several of these "hard core" subroutines mentioned above.

Interval arithmetic enables a subroutine user to *rigorously* state that the result of a calculation is correct to a certain number of decimal places. Generally, this has not

been possible with a computer before the development of the concept of interval arithmetic.

The abstracts themselves are divided into three main groups:

Group 1.

General computational procedures from elementary mathematics with no error bounding.

Group 2.

The Interval Arithmetic System and its subroutines.

Group 3.

General computational procedures from elementary mathematics with error bounding. All the subroutines in this group use the Interval Arithmetic System of Group 2.

Only subroutines in Group 1 will appear in this article. The routines from Groups 2 and 3, as well as the nonlinear least squares routines, will appear in a future SPS.

The presentation of the abstracts will follow according to this format:

a. Identification

- (1) Program or subroutine name
- (2) Source language
- (3) Machine requirements

b. Purpose of the subroutine.

- c. Mathematical methods used in the numerical procedure, together with appropriate references.
- d. Computational experience with the subroutine, miscellaneous information, and credits.

1. Group 1: Abstract 1

a. Identification

- (1) Subroutine name: MATIN2
- (2) Source language: FORTRAN IV
- (3) Machine: Any for which a FORTRAN IV compiler is available.

b. Purpose. Let $A = \{a_{ij}\}$ and $b = \{b_{ij}\}$, respectively, represent $n \times n$ and $n \times m$ real matrices. The subroutine MATIN2 attempts to approximate the solution to the matrix equation $Ax = b$, provided this solution exists.

Matrix inversions, solutions of linear equations, and determinant evaluations are possible with MATIN2. The accuracy of the calculations is single precision.

c. Mathematical method. Jordan's method (Ref. 1) is used to invert the matrix A and to obtain $x = A^{-1}b$. Full pivoting is used.

d. Experience. Computational experience with MATIN2 has shown it to be free of coding errors. For badly conditioned matrices A , the solution returned by MATIN2 may have no correct digits at all. Thus this routine should be used with extreme caution.

The present program MATIN2 is a slight modification of an existing JPL subroutine, which was a modification of SHARE program 664, contributed by Argonne National Laboratory.

2. Group 1: Abstract 2

a. Identification

- (1) Subroutine name DINVR2

- (2) Source language FORTRAN IV

- (3) See abstract 1, item a(3)

b. } See abstract 1, items b and c, and Ref. 1.
c. }

d. Experience. This subroutine is a double precision version of the subroutine MATIN2 presented in abstract 1.

3. Group 1: Abstract 3

a. Identification

- (1) Subroutine name: SOLVE
- (2) Source language: FORTRAN IV-MAP
- (3) Machine: IBM 7094

b. Purpose. Let $A = \{a_{ij}\}$ and $b = \{b_i\}$, respectively, denote an $n \times n$ nonsingular real matrix and an n -dimensional real vector. This routine obtains a highly accurate approximation to the solution of the system $Ax = b$. Usually the true solution (correctly rounded to fit an IBM 7090/94 word) can be reached.

The routine also has the capability of obtaining solutions for many vectors b without further processing of the matrix $A = \{a_{ij}\}$.

c. Mathematical method. The system is first scaled so that the element of maximum modulus of each row lies between -1 and $-\frac{1}{2}$ or $\frac{1}{2}$ and 1 .

The first entry to SOLVE then factors A into the product of a lower triangular matrix L and an upper triangular matrix U . Thus the system $Ax = b$ becomes the system $LUx = b$. This latter system is equivalent to the two triangular systems $Lw = b$ and $Ux = w$, both of which are easily solved.

A first approximation, x_0 , is thus obtained. The residual vector $r_0 = b - Ax_0$ is calculated by forming double precision inner products in the components of Ax_0 with a final rounding of $b - Ax_0$ to single precision.

We then set $x_1 = x_0 + dx_0$ and solve the system $Adx_0 = LUdx_0 = r_0$ for dx_0 by solving the systems $Ldw = r_0$ and $Udx_0 = dw$.

We now form $r_1 = b - Ax_1$ and repeat the same procedure as above to form a sequence of vectors x_0, x_1, x_2, \dots until a certain member of this sequence has a specified relative accuracy.

If x_0 is accurate to s digits, then dx_0 will be accurate to s digits. Thus x_1 is (roughly) accurate to $2s$ digits; x_2 is accurate to $3s$ digits, etc.

Note that if $s = 0$, no improvement in accuracy may be made at each iteration. This situation may occur when A is poorly conditioned.

It should be emphasized that the great increase in accuracy which is usually possible with SOLVE is due primarily to the technique employed in computing the residual vectors $r_i = b - Ax_i$, ($i = 0, 1, \dots$).

See Refs. 2-4 for further details.

d. Experience. Computational experience with SOLVE has shown it to be free of coding errors.

For badly conditioned matrices A , the solution may have no correct digits; this will, however, usually be noted by the program.

This routine should be used with some caution. The user should pay close attention to the program's decisions regarding a singularity or failure of the iterative improvement for the matrix A .

This program is a FORTRAN IV modification of SHARE program SDA 3194.

4. Group 1: Abstract 4

a. Identification

- (1) Subroutine name: SLVINV
- (2) Source language: FORTRAN IV
- (3) Machine: IBM 7094

b. Purpose. This routine obtains a highly accurate approximation to the inverse of an $n \times n$ real nonsingular matrix $A = \{a_{ij}\}$.

c. Mathematical method. This routine solves the n linear systems $Ax_i = e_i$, ($i = 1, \dots, n$), where the e_i are the columns of the $n \times n$ identity matrix, using the routine SOLVE of abstract 3.

The vectors x_i are the columns of a right approximate inverse for A .

If a left approximate inverse is desired for A , one can solve the n systems $A^T y_i = e_i$, ($i = 1, \dots, n$). The row vectors y_i^T , ($i = 1, \dots, n$), form the consecutive rows of

a left approximate inverse. This can be effected with the routine SLVINV by forming the transpose of a right approximate inverse for A^T .

d. Experience. Computational experience with SLVINV has shown it to be free of coding errors.

This routine is the most reliable and accurate single precision matrix inversion program available. It should, however, still be used with some caution.

5. Group 1: Abstract 5

a. Identification

- (1) Subroutine name: DSOLVE
- (2) Source language: FORTRAN IV-MAP
- (3) Machine: IBM 7094

b. } See abstract 3, items b and c, and Refs. 2-4.
c. }

d. Experience. This routine is a double precision version of the routine SOLVE presented in abstract 3. Experience with DSOLVE has shown it to be free of coding errors.

The residual vectors $r_i = b - Ax_i$ (see abstract 3) are calculated using quadruple precision accumulated inner products with a final rounding to double precision on all components of the vector $b - Ax_i$, ($i = 0, 1, 2, \dots$).

If one desires an accurate approximation for the inverse of the matrix $A = \{a_{ij}\}$, the vectors x_i , ($i = 1, \dots, n$), obtained by solving the n linear systems $Ax_i = e_i$ (see abstract 4) will form consecutive columns for an approximate right inverse.

6. Group 1: Abstract 6

a. Identification

- (1) Subroutine name: LSQSOL
- (2) Source language: FORTRAN IV-MAP
- (3) Machine: IBM 7094

b. Purpose. Let $A = \{a_{ij}\}$ and $b = \{b_i\}$ respectively represent a real $m \times n$ matrix and a real m -dimensional vector. No assumptions are made regarding either the rank of the matrix A or the relations of the positive integers m and n to each other.

This routine will approximate the unique solution of minimal euclidean length (the pseudoinverse solution)

for the least squares problem $Ax = b$, as well as generate an orthonormal basis for the null-space of A , $\{x | Ax = 0\}$.

The routine has the capability of processing many vectors b without further processing of $A = \{a_{ij}\}$.

c. Mathematical method. Basic mathematical algorithm. Let $r_1 = \min(m, n)$ and $r = \text{rank } A$. An $m \times m$ orthogonal matrix Q is constructed as a product of Householder reflections (Refs. 5 and 6), such that

$$QA = \begin{matrix} & \begin{matrix} \overbrace{r} & \overbrace{n-r} \end{matrix} \\ \begin{matrix} r \\ m-r \end{matrix} & \left\{ \begin{bmatrix} \overbrace{R_{11}} & \overbrace{R_{12}} \\ 0 & 0 \end{bmatrix} \right\} \end{matrix} \begin{matrix} \\ m \end{matrix} \quad (1)$$

The matrix R_{11} is a nonsingular upper triangular matrix.

If $r < n$, we construct an $n \times n$ orthogonal matrix S , again as a product of Householder 'reflections' (Ref. 7), such that

$$QAS = \begin{matrix} & \begin{matrix} \overbrace{r} & \overbrace{n-r} \end{matrix} \\ \begin{matrix} r \\ m-r \end{matrix} & \left\{ \begin{bmatrix} \overbrace{R} & \overbrace{0} \\ 0 & 0 \end{bmatrix} \right\} \end{matrix} \quad (2)$$

The matrix R is a nonsingular $r \times r$ upper triangular matrix. [In case $r = r_1$, the matrix displayed in the right member of Eq. (2) must be properly interpreted.]

Let

$$Qb = \begin{bmatrix} c \\ d \end{bmatrix} \begin{matrix} r \\ m-r \end{matrix} \quad (3)$$

$$x = Sy \quad (4)$$

and

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \begin{matrix} r \\ n-r \end{matrix} \quad (5)$$

Since the length of an m -dimensional vector is invariant under left multiplications by the $m \times m$ orthogonal matrix Q , the euclidian length of $Ax - b$ is minimized precisely when

$$[R \ 0] y = Ry_1 = c \quad (6)$$

This system of r equations in Eq. (6) can be easily solved since R is upper triangular.

The vector y of Eq. (5) has only the segment y_1 uniquely determined by Eq. (6). The segment y_2 can be chosen arbitrarily.

The unique solution of Eq. (6) of minimal euclidean length, however, is the vector

$$y = \begin{bmatrix} y_1 \\ 0 \end{bmatrix} = \begin{bmatrix} R^{-1}c \\ 0 \end{bmatrix} \quad (7)$$

for which $y_2 = 0$.

Thus, since S is orthogonal, the unique least squares solution of $Ax = b$ of minimal euclidean length is given by $x = Sy$, where y is given by Eq. (7).

An orthonormal basis for the null space of the matrix QAS of Eq. (2) is given by the columns of the matrix

$$X = \begin{bmatrix} \overbrace{0}^{n-r} \\ I \end{bmatrix} \begin{matrix} r \\ n-r \end{matrix} \quad (8)$$

Here I is the $n - r \times n - r$ identity matrix.

Again using the fact that S is orthogonal, the $n \times n - r$ matrix

$$H = SX \quad (9)$$

has as its columns an orthonormal basis for the null space of A , $\{x | Ax = 0\}$.

d. Numerical algorithm. Instead of r , the actual rank of A , which cannot, in general, be calculated with a fixed precision of arithmetic for a general matrix, we calculate a pseudorank or (ϵ) rank which is defined in the following way.

Let $\epsilon > 0$ be given. Suppose that $0 \leq k < r_1$ steps of the forward triangularization procedure have been performed so that

$$Q_k A = k \left\{ \begin{bmatrix} \overbrace{R_{11}}^k & \overbrace{R_{12}}^{n-k} \\ 0 & \epsilon R_{22} \end{bmatrix} \right\} \begin{matrix} m-k \end{matrix} \quad (10)$$

Here the matrix R_{11} is a $k \times k$ nonsingular upper triangular matrix.

Let $R_{12} = \{r_{ij}\}$, ($j = k + 1, \dots, n$), ($i = 1, \dots, k$), and $R_{22} = \{t_{ij}\}$, ($j = k + 1, \dots, n$), ($i = k + 1, \dots, m$).

We then define the (ϵ) rank A to be the smallest nonnegative integer k for which

$$\sum_{i=k+1}^m t_{ij}^2 \leq \sum_{i=1}^k r_{ij}^2 \quad (j = k+1, \dots, n) \quad (11)$$

In particular, the following theorem may be proved.

Theorem. If A has (ϵ) rank k for a given $\epsilon > 0$, then exactly $n - k$ of the nonnegative eigenvalues of $A^T A$ do not exceed $\epsilon^2 \|R_{22}^T R_{22}\|$, where R_{22} is given in Eq. (10). (The matrix norm indicated here is the euclidean norm.)

If a given matrix A has (ϵ) rank $r < r_1$ for a given $\epsilon > 0$, we next identify the matrix in the right member of Eq. (10) with the matrix in the right member of Eq. (1). We then compute the solutions of minimal euclidean length for this new system.

Iterative improvements are calculated by noting that if x_0 is an approximate solution, and if $x_1 = x_0 + dx_0$, then $Ax_1 - b = Adx_0 - r_0$, where $r_0 = b - Ax_0$. This is a new least squares problem for dx_0 .

In this way one obtains a sequence of vectors x_0, x_1, \dots ; this sequence will terminate when some member of it has a specified relative accuracy. Let us call the solution so obtained x' .

In case $r < n$, we accept this as a solution to the least squares problem.

If $r = n$, we observe that since Q is orthogonal the system

$$A^T Ax = A^T b \quad (12)$$

is the system

$$R^T Rx = A^T b \quad (13)$$

where R is given in Eq. (2).

The system in Eq. (13) is equivalent to the two $n \times n$ triangular systems

$$R^T w = A^T b \quad (14)$$

and

$$Rx = w \quad (15)$$

Eqs. (14) and (15) are easily solved.

We again calculate iterative refinements as above to obtain a second sequence of vectors $\tilde{x}_0, \tilde{x}_1, \dots$. We thus obtain a certain member, say x'' , of this sequence which has a specified relative accuracy.

Double precision inner products are accumulated in forming the vectors $r_i = b - Ax_i$ and $A^T(b - A\tilde{x}_i)$, ($i = 0, 1, \dots$).

The solution returned is the vector x' or x'' for which $\|Ax - b\|$ is minimized. Roughly speaking, a problem $Ax - b$ for which $Ax - b = 0$ will usually result in $x = x'$. On the other hand, problems for which $Ax - b \neq 0$ often result in $x = x''$.

Experience has shown that LSQSOL is free of coding errors and it is highly recommended.

The accuracy of the computation is essentially single precision, but the solution returned is in double precision. The number of correct digits in this solution may be more than that allowed by single precision representation.

7. Group 1: Abstract 7

a. Identification

- (1) Subroutine name: COVLSQ
- (2) Source language: FORTRAN IV
- (3) Machine: IBM 7094

b. Purpose. Let $A = \{a_{ij}\}$ be an $m \times n$ real matrix of rank n . Then COVLSQ calculates $(A^T A)^{-1}$ (without forming $A^T A$) with output from the routine LSQSOL of abstract 6.

c. Mathematical method

Since $QA = \begin{bmatrix} R \\ 0 \end{bmatrix}$, where R is the nonsingular upper triangular $n \times n$ matrix defined in abstract 3, Eq. (2), we have

$$A^T A = (QA)^T QA = R^T R \quad (1)$$

Thus

$$(A^T A)^{-1} = R^{-1} (R^{-1})^T \quad (2)$$

Since R is upper triangular, R^{-1} is also. Hence $(A^T A)^{-1}$ is obtained as the product of an upper triangular and a lower triangular matrix.

d. Experience. This routine can be expected to retain more accuracy (with single precision calculations) than a routine which would invert $A^T A$ directly.

Experience has shown this routine to be free of coding errors.

If one has solved a least squares problem $Ax = b$ with rank $A = n$, and if $m > n$, then the variance-covariance matrix is given by $[\sigma^2/(m-n)] (A^T A)^{-1}$, where $\sigma^2 = \|Ax - b\|^2$.

8. Group 1: Abstract 8

a. Identification

- (1) Subroutine name: SEQLSQ/SEQLQ2
- (2) Source language: FORTRAN IV-MAP
- (3) Machine: IBM 7094

b. Purpose. This routine obtains the least squares solution of an arbitrarily dimensioned linear least squares system $Ax = b$ where, because of lack of storage or non-availability of the data, not all of the rows of the $k \times n$ coefficient matrix $A = \{a_{ij}\}$ can be processed at each entry into SEQLSQ.

c. Mathematical method. Suppose that the linear least squares problem

$$A_1 x_1 = b_1 \quad (1)$$

is given, where A_1 is a $k_1 \times n$ real matrix and b_1 is a k_1 -dimensional vector.

Let us assume that the least squares problem of Eq. (1) is changed so that we have the new least squares problem

$$\begin{bmatrix} A_1 \\ A_2 \end{bmatrix} x_2 = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \quad (2)$$

where A_1 and b_1 are as in Eq. (1) while A_2 is a $k_2 \times n$ real matrix and b_2 is a k_2 -dimensional vector.

Solutions for both Eqs. (1) and (2) may be obtained with the methods presented in abstract 6 of the routine LSQSOL.

Frequently one considers a problem

$$\begin{bmatrix} A_1 \\ \vdots \\ A_m \end{bmatrix} x_m = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \quad (3)$$

where each A_i , ($i = 1, \dots, m$), is a $k_i \times n$ real matrix and b_i is a k_i -dimensional vector. The number

$$k = \sum_{i=1}^m k_i$$

is often quite large; typically the integer kn exceeds the machine's storage capacity. The significant point here is that one may solve the least squares problem of Eq. (3) using no more than $(k_{max} + n + 1)(n + 1)$ machine cells, where

$$k_{max} = \max_{1 \leq i \leq m} k_i.$$

Not all of the data for the matrices A_i and the vectors b_i , ($i = 1, \dots, m$), need be available simultaneously.

For example, if $n = 6$, $m = 30,000$, and $k_i = 2$, ($i = 1, \dots, m$), only $(2 + 6 + 1)(6 + 1) = 63$ machine cells need be used.

One may also calculate the inverse of the matrix $A^T A$ where A is the matrix in the left member of Eq. (3). This is accomplished by the method presented in abstract 8 for the routine COVLSQ.

d. Experience. The routine SEQLSQ/SEQLQ2 has been used and appears to be free of coding errors. The accuracy of the computation is double precision.

In particular, it can be shown that for a given problem such as Eq. (3) where the condition number of the coefficient matrix exceeds 10^8 , there are vectors b_i , ($i = 1, \dots, m$), for which SEQLSQ/SEQLQ2 obtains a least squares solution accurate to 8 decimal places but which would require quadruple precision in forming and solving the normal equations associated with Eq. (3) to obtain 8 decimal digits of accuracy in the solution.

9. Group 1: Abstract 9

a. Identification

- (1) Subroutine name: SYMEIG
- (2) Source language: FORTRAN IV
- (3) Machine: IBM 7094

b. Purpose. Let $A = \{a_{ij}\}$ be an $n \times n$ real symmetric matrix. This routine calculates the eigenvalues and eigenvectors of A .

c. Mathematical method. An orthogonal matrix Q is constructed as a product of Householder 'reflections' such that the matrix $H = QAQ^T$ is tridiagonal. The eigenvalues of H , which are, of course, also those of A , are calculated using bisectioning together with Sturm sequence properties of successive principal minors of the matrix H (Refs. 8 and 9). Eigenvectors $\{w_i\}$ of H are calculated using the inverse power method (see Ref. 10) so that the eigenvectors $\{v_i\}$ of A are given by $v_i = Q^T w_i$, ($i = 1, \dots, n$).

d. Experience. This routine appears to be free of coding errors.

The accuracy is essentially single precision.

If A has a multiple eigenvalue λ_j , a full set of eigenvectors for A corresponding to the eigenvalue λ_j will not be calculated by SYMEIG, since equal eigenvalues return equal eigenvectors. This difficulty may be overcome by obtaining an orthonormal basis for the null space of the matrix $A - \lambda_j I$ with the routine LSQSOL of abstract 6. Here I designates the $n \times n$ identity matrix.

This routine is a modification of SHARE program SDA3202-01.

10. Group 1: Abstract 10

a. Identification

- (1) Subroutine name: QREIG
- (2) Source language: FORTRAN IV
- (3) Machine: IBM 7094

b. Purpose. Let $A = \{a_{ij}\}$ designate an $n \times n$ real matrix. This routine calculates the real and complex eigenvalues of A .

c. Mathematical method. The matrix A is first transformed to upper Hessenberg form (Ref. 11) with a sequence of similarity transformations. The eigenvalues of this upper Hessenberg matrix are calculated using the Q-R transform method of J. G. F. Francis (Ref. 12).

d. Experience. The computation is performed essentially in single precision.

Experience with QREIG has shown it to be free of coding errors. Frequently the eigenvalues returned by QREIG may have only a few digits correct. This situation may be improved with the routine EVCOR of

abstract 11, which corrects the eigenvalues and obtains eigenvectors for diagonalable matrices with real or complex eigenvalues.

This program is a modification of SHARE program SDA 3006.

11. Group 1: Abstract 11

a. Identification

- (1) Subroutine name: EVCOR
- (2) Source language: FORTRAN IV-MAP
- (3) Machine: IBM 7094

b. Purpose. This routine improves the accuracy of approximate eigenvalues of real diagonalable matrices. These eigenvalues can be either real or complex.

c. Mathematical method. Let λ^* be an approximation to an eigenvalue λ_i of a real $n \times n$ matrix $A = \{a_{ij}\}$. Suppose that $\lambda^* = \lambda_i - e_i$ where $|e_i| > 0$, and that $|\lambda_j - \lambda^*| > |e_i|$ for any other eigenvalue λ_j of A which is distinct from λ_i .

Since A is diagonalable, the complex n -dimensional coordinate space C_n is spanned by a set of n linearly independent eigenvectors $\{u_i\}$ of A .

The vector

$$v_0 = \begin{cases} (1, \dots, n), \lambda^* \text{ real} \\ (1 + (n+1)i, \dots, n + 2ni), \lambda^* \text{ complex} \end{cases} \quad (1)$$

is taken as the starting vector of the iteration

$$(A - \lambda^* I) v_{r+1} = v_r, (r = 0, 1, 2, \dots) \quad (2)$$

This is the inverse power method.

If

$$v_0 = \sum_{k=1}^n \beta_k u_k,$$

then

$$v_r = e_i^{-r} \left[v + \sum_{\substack{k=1 \\ \lambda_k \neq \lambda_i}}^n \beta_k \left(\frac{e_i}{\lambda_k - \lambda^*} \right)^r u_k \right]. \quad (3)$$

The vector v appearing in Eq. (3) is a linear combination of (possibly several) eigenvectors of A corresponding to the eigenvalue λ_i .

From the fact that $|e_i/(\lambda_k - \lambda^*)| < 1$ for all eigenvalues $\lambda_k \neq \lambda_i$, v_r will round (in an IBM 7094 word, say) to the eigenvector $u = e_i^{-r} v$ for sufficiently large r .

Let u_m be a component of the vector u of largest magnitude, and let a_1, \dots, a_n denote the consecutive rows of the matrix A . Then since $\lambda_i u = Au$, we have λ_i as the complex inner product $\lambda_i = (a_m, u)/u_m$. See Ref. 10 for further details.

d. Experience. The linear systems in Eq. (2) are solved with the routine SOLVE presented in abstract 3.

This routine appears to be free of coding errors. The accuracy is essentially single precision. The hypotheses needed to effectively use EVCOR are critical.

If A has distinct real or complex eigenvalues, or is symmetric, then EVCOR may be used.

12. Group 1: Abstract 12

a. Identification

- (1) Subroutine name: POLYRT
- (2) Source language: FORTRAN IV
- (3) Machine: IBM 7094

b. Purpose. This routine obtains the real and complex roots of an n th degree polynomial

$$P(x) = \sum_{i=0}^n a_i x^i$$

with real coefficients.

c. Mathematical method. This routine uses a combined Newton-Muller method. The root is located "roughly" with the Muller method and then "refined" with Newton's method. Muller's method is defined as follows:

Given 3 distinct complex points z_1, z_2 , and z_3 one interpolates $P(z_1)$, $P(z_2)$ and $P(z_3)$ with a quadratic; this quadratic is then solved to obtain a new point z_4 . The process begins anew with the points z_2, z_3 and z_4 , and continues until $|P(z)|$ is "small."

The Newton method is defined as the sequence (Refs. 13 and 14)

$$z_{n+1} = z_n - \frac{P(z_n)}{dP(z_n)/dz}$$

$$(n = 1, 2, \dots).$$

d. Experience. The accuracy of the computation is double precision.

This program appears to be free of coding errors; it may fail for polynomials with multiple roots. This routine is a modification of SHARE program SDA 3332.

References

1. Fox, L., *An Introduction to Numerical Linear Algebra*, pp. 65-75, Oxford University Press, New York, 1964.
2. Wilkinson, J. H., *Rounding Errors in Algebraic Processes*, pp. 121-126, Prentice-Hall, Inc., New York, 1963.
3. Forsythe, G., and Moler, C., *Computer Solutions of Linear Algebraic Systems*, Prentice-Hall, Inc., New York, 1967.
4. Moler, C., "Iterative Refinement in Floating Point." *J. Assoc. Comp. Mach.*, Vol. 14, No. 2, pp. 316-321, New York, April, 1967.
5. Householder, A. S., "Unitary Triangularization of a Nonsymmetric Matrix," *J. Assoc. Comp.*, pp. 339-342, March 5, 1958.
6. Businger, P., and Golub, G., "Linear Least Squares Solutions by Householder Transformations," *Num. Math.*, Vol. 7, pp. 269-276, Berlin-Wilmersdorf, Germany, 1965.
7. Lawson, C. L., and Hanson, R. J., "Extensions of the Golub-Householder Algorithm for Solving Linear Least Squares Problems" (to be published).
8. Wilkinson, J. H., *The Algebraic Eigenvalue Problem*, Oxford University Press, New York, 1965.
9. *A Survey of Numerical Analysis*, pp. 245-247. Edited by J. Todd. McGraw-Hill Book Co., Inc., New York, 1962.
10. Wilkinson, J. H., *The Algebraic Eigenvalue Problem*, pp. 229-344, Clarendon Press, Oxford, England, 1965.
11. Fox, L., *An Introduction to Numerical Linear Algebra*, p. 364, Oxford University Press, New York, 1965.
12. Francis, J. G. F., "The Q-R Transformation, Parts I and II," *Comput. J.*, Vol. 4, pp. 265-271, 332-345, 1961.
13. Muller, D. E., "A Method for Solving Algebraic Equations Using an Automatic Computer," *Mathematical Tables and Other Aids to Computations*, Vol. 10, p. 208, 1956.
14. Nielsen, K. J., *Methods in Numerical Analysis*, p. 139, The Macmillan Company, New York, 1956.

B. A Numerical Integration of Lunar Motion Employing a Consistent Set of Constants, C. J. Devine

1. Introduction

In SPS 37-47, Vol. III, pp. 8-19, Devine and Lawson wrote an article describing a Newtonian integration and fit to the lunar ephemeris (LE4). That integration employed constants not necessarily consistent with the current constants preferred at JPL or consistent with any

available lunar ephemeris. This article describes an integration and fit to the lunar ephemeris LE4 (Ref. 1), using consistent constants provided by Mulholland in a recent article (SPS 37-47, Vol. III, pp. 6-7). In addition, the integration was performed over an extended period (2 yr) in order to provide a comparison to *Lunar Orbiter* ranging data and *Surveyor* data¹ (Ref. 2). The fitted ephemeris described by this article is now designated as LE5.²

2. Numerical Integration and Fit of the Moon

A numerical integration and least squares fit using PLOD II (Ref. 3) were made to LE4 over a 2-yr period (732 days) from JD 243 9240.5 (April 25, 1966) to JD 243 9972.5 (April 26, 1968) using a set of constants consistent with the evaluation of the Brown Lunar Theory used in the computation of LE4. Maximum difference in the rectangular coordinates between the fitted ephemeris and LE4 in the sense of PLOD II - LE4 was 543 m, as given in Table 1. The integrator used a step size of 1/4 day and carried backward differences of the acceleration

¹Sjogren, W. L., "Lunar Orbiter Ranging Residuals Using LE4 Versus Range Residuals From PLOD Integration and LE4," JPL Sect. 311 interoffice memorandum, 311.1-24, Sept. 26, 1967.

²Mulholland, J. D., "Announcement of JPL Development Ephemeris No. 29," JPL Sect. 314 interoffice memorandum, 314.13-52, Oct. 5, 1967.

Table 1. Maximum residuals for PLOD II fit to LE4, over the interval JD 243 9240.5 (April 25, 1966) to JD 243 9972.5 (April 26, 1968)

Maximum position residuals	
$\delta x = 0.3628 \times 10^{-8}$ AU = 543 m	
$\delta y = 0.2797 \times 10^{-8}$ AU = 418 m	
$\delta z = 0.2239 \times 10^{-8}$ AU = 335 m	
Maximum polar residuals	
$\delta R/R = 136.7 \times 10^{-8}$	
$\cos \beta \delta \lambda = 78.91 \times 10^{-8}$ rad	
$\delta \beta = 58.81 \times 10^{-8}$ rad	
Maximum polar residuals multiplied by mean distance	
$\bar{R} = 2.570 \times 10^{-2}$ AU	
$\bar{R} \cdot \delta R/R = 0.3513 \times 10^{-8}$ AU = 525.5 m	
$\bar{R} \cdot \cos \beta \delta \lambda = 0.2028 \times 10^{-8}$ AU = 303.0 m	
$\bar{R} \cdot \delta \beta = 0.1511 \times 10^{-8}$ AU = 226.0 m	
Maximum velocity residuals	
$\delta \dot{x} = 0.1240 \times 10^{-8}$ AU/day = 2.15 mm/s	
$\delta \dot{y} = 0.1170 \times 10^{-8}$ AU/day = 2.03 mm/s	
$\delta \dot{z} = 0.07816 \times 10^{-8}$ AU/day = 1.35 mm/s	

through the 14th order using the predictor-corrector option at each step. The fitted lunar rectangular coordinates were geocentric 1950.0 equatorial and the residuals of the coordinates were expressed in AU and the velocity residuals were expressed in AU/day. The residuals were also presented in 1950.0 ecliptic polar coordinates: geocentric radius R , geocentric longitude λ , and geocentric latitude β . The graphs of the residuals appear in Figs. 1-3.

3. Parameters Solved

Changes to the value of the lunar ephemeris scale factor REM, and the GM of the barycenter were determined to be unnecessary with the new value of the GM of the earth and the value of the AU in kilometers (Table 2)

Table 2. Constants used in the numerical integration and fit to the lunar ephemeris LE4

Planet	GM AU ³ /s ²	GM km ³ /s ²
Mercury	0.4931870138093182 D-10	22118.75300341569
Venus	0.7252750203078209 D-09	325275.7794619955
Earth	0.8887706500323706 D-09	398601.300000000
Mars	0.9565612034446127 D-10	42900.44222417784
Jupiter	0.2825328644877725 D-06	126712068.0385296
Saturn	0.8450771312702505 D-07	37900536.33210365
Uranus	0.1293944677448034 D-07	5803162.272967516
Neptune	0.1532112500184275 D-07	6871311.899166103
Pluto	0.8219783563488637 D-09	368645.8833902616
AU = 149597900.0 km		EM ratio = 81.302
K = 0.017202098950		J2 = 0.11115700 E-02

Table 3. Reciprocal masses used in the integration and fit to the lunar ephemeris LE4

Planet	M ⁻¹
Moon	27069136.85
Mercury	6000000.0
Venus	408000.0
Earth	332945.52
Mars	3093500.0
Jupiter	1047.355
Saturn	3501.6
Uranus	22869.0
Neptune	19314.0
Pluto	360000.0
Barycenter	328900.11

input to the program. With these values held constant, changes to the six parameters x_0, y_0, z_0 and $\dot{x}_0, \dot{y}_0, \dot{z}_0$ at the initial epoch of the integration were computed using PLOD II. The partial derivatives needed for this computation were computed by PLOD II using two-body conic approximations. This procedure was iterated until there were no further corrections to be made. The GM of the earth, the AU in kilometers, the earth-moon mass ratio, the value of J2 (Table 2), and the inverse masses of the other eight planets (Table 3) were input to the program. The GM of the moon and the barycenter (Table 4) were computed from the above.

The final values determined for these nine parameters are given in Table 4. Included also are the values of the osculating conic elements a_0, n_0, e_0 , and M_0 , the semi-major axis, mean motion, eccentricity and mean anomaly at epoch, which were computed from the rectangular coordinates at epoch.

Table 4. Final epoch values determined for the parameters $x_0, y_0, z_0, \dot{x}_0, \dot{y}_0, \dot{z}_0$, and REM, and the corresponding elements a_0, n_0, e_0 , and M_0 for the PLOD II integration and fit to LE4

Epoch values equatorial JD 243 9240.5
$x_0 = 2.778803790115670 \text{ D} - 04, \text{ AU}$
$y_0 = 2.272984042933926 \text{ D} - 03, \text{ AU}$
$z_0 = 1.107378914796110 \text{ D} - 03, \text{ AU}$
$\dot{x}_0 = -5.933118703149390 \text{ D} - 04, \text{ AU/day}$
$\dot{y}_0 = 1.933768789802169 \text{ D} - 05, \text{ AU/day}$
$\dot{z}_0 = 5.953404392620535 \text{ D} - 05, \text{ AU/day}$
$a_0 = 2.559726047434522 \text{ D} - 03, \text{ AU}$
$n_0 = 2.316110099979990 \text{ D} - 01, \text{ rad/day}$
$e_0 = 3.677771507368935 \text{ D} - 02$
$M_0 = 4.920640450425223 \text{ D} + 00, \text{ rad}$
REM = 6378.1495 km
$GM_{\oplus} = 0.1093171939229503 \text{ D} - 10, \text{ AU}^3/\text{day}^2$
$= 4902.724410 \text{ km}^3/\text{s}^2$
$GM_{\oplus} + GM_{\text{m}} = 0.8997023694246655 \text{ D} - 09, \text{ AU}^3/\text{day}^2$
$= 403504.0244 \text{ km}^3/\text{s}^2$

4. Conclusion

The work reported here will be used as an experimental lunar ephemeris. The interpretation of the integration and fit will be described by Mulholland in SPS 37-49, Vol. III, and in Ref. 4. The use of the appropriate constants has provided an integration of the lunar ephemeris which has produced residuals between *Lunar Orbiter* ranging data and the integrated ephemeris which are substantially reduced from those computed with LE4. In addition, certain gravitational anomalies in LE4 have become apparent. The numerical solution of the set of Newtonian differential equations included a J2 term (Table 2) in the Earth's potential function and were otherwise based on a point mass model. Perturbations by all planets were included. The partial derivatives computed by PLOD II using two-body conic approximations (Ref. 5, p. 241) were adequate for the time period involved, but convergence was slow. It is recommended that whenever a longer time period is involved in future lunar integration, that a different differential correction procedure be utilized.

In addition the lunar ephemeris has been numerically integrated by the double precision orbit determination program³ with similar results, over a smaller period (1 mo). It is significant that this procedure also demonstrates that the integrated lunar ephemeris is basically more representative of the actual motion of the moon.

References

1. Mulholland, J. D., and Block, N., "JPL Lunar Ephemeris No. 4," Technical Memorandum 33-346, Jet Propulsion Laboratory, Pasadena, Calif., July 15, 1967.
2. Cary, C. N., and Sjogren, W. L., "Lunar Ephemeris Errors Confirmed by Radio Observations of Lunar Probes," to be presented to American Astronomical Society Meeting, University of Pennsylvania, Philadelphia, Penn.
3. Devine, C. J., "PLOD II: Planetary Orbit Determination Program for the IBM 7094 Computer," Technical Memorandum 33-188, Jet Propulsion Laboratory, Pasadena, Calif., April 15, 1965.
4. Mulholland, J. D., Devine, C. J., "Gravitational Inconsistency in the Lunar Ephemeris," (to be published externally).
5. Brouwer, D., and Clemence, G. M., "Methods of Celestial Mechanics," Academic Press, New York, N. Y., 1961.

³Sturms, F. M., "An Integrated Lunar Ephemeris," JPL Section 312 internal memorandum 312-848, Sept. 22, 1967.

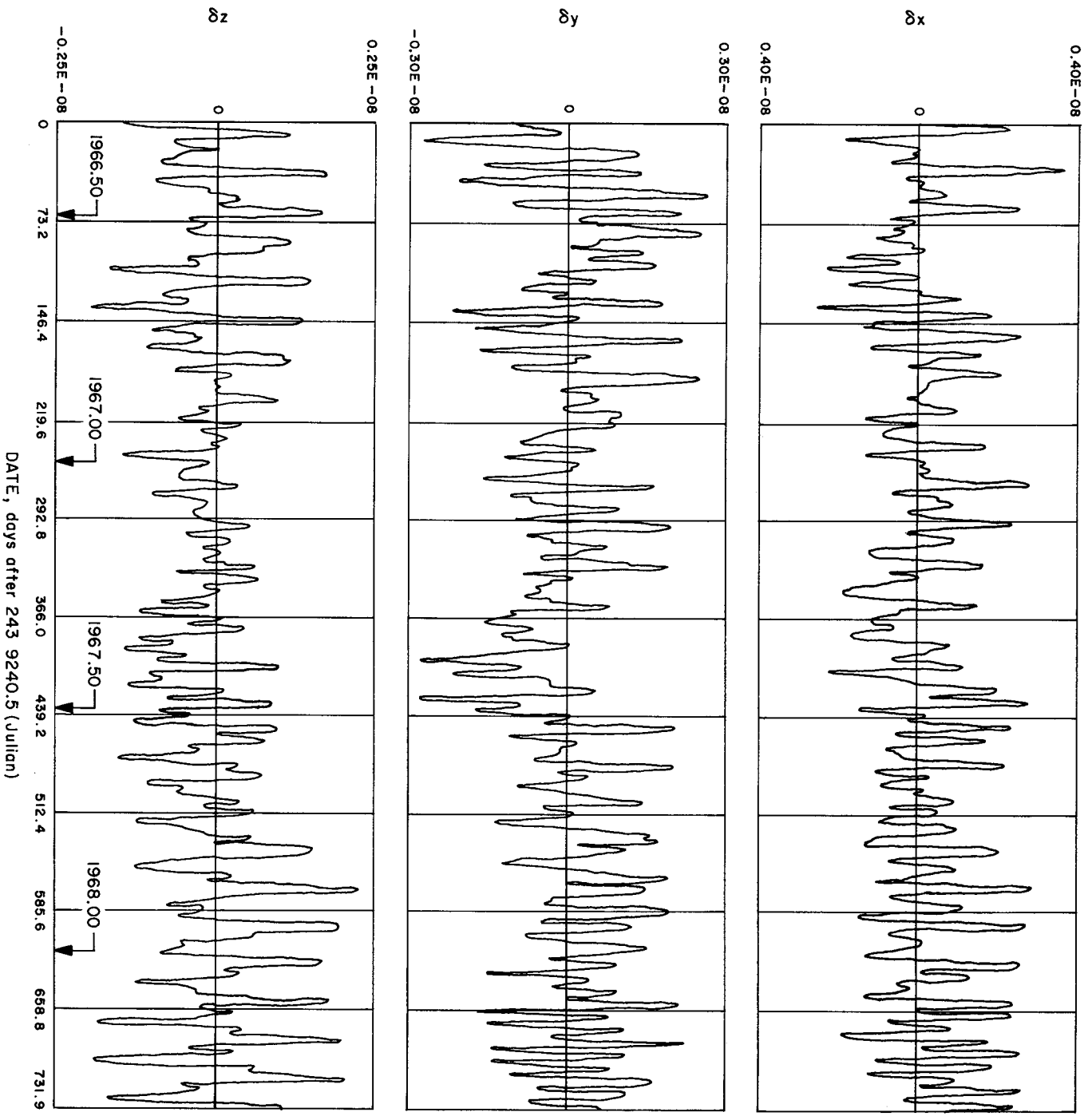


Fig. 1. Rectangular residuals

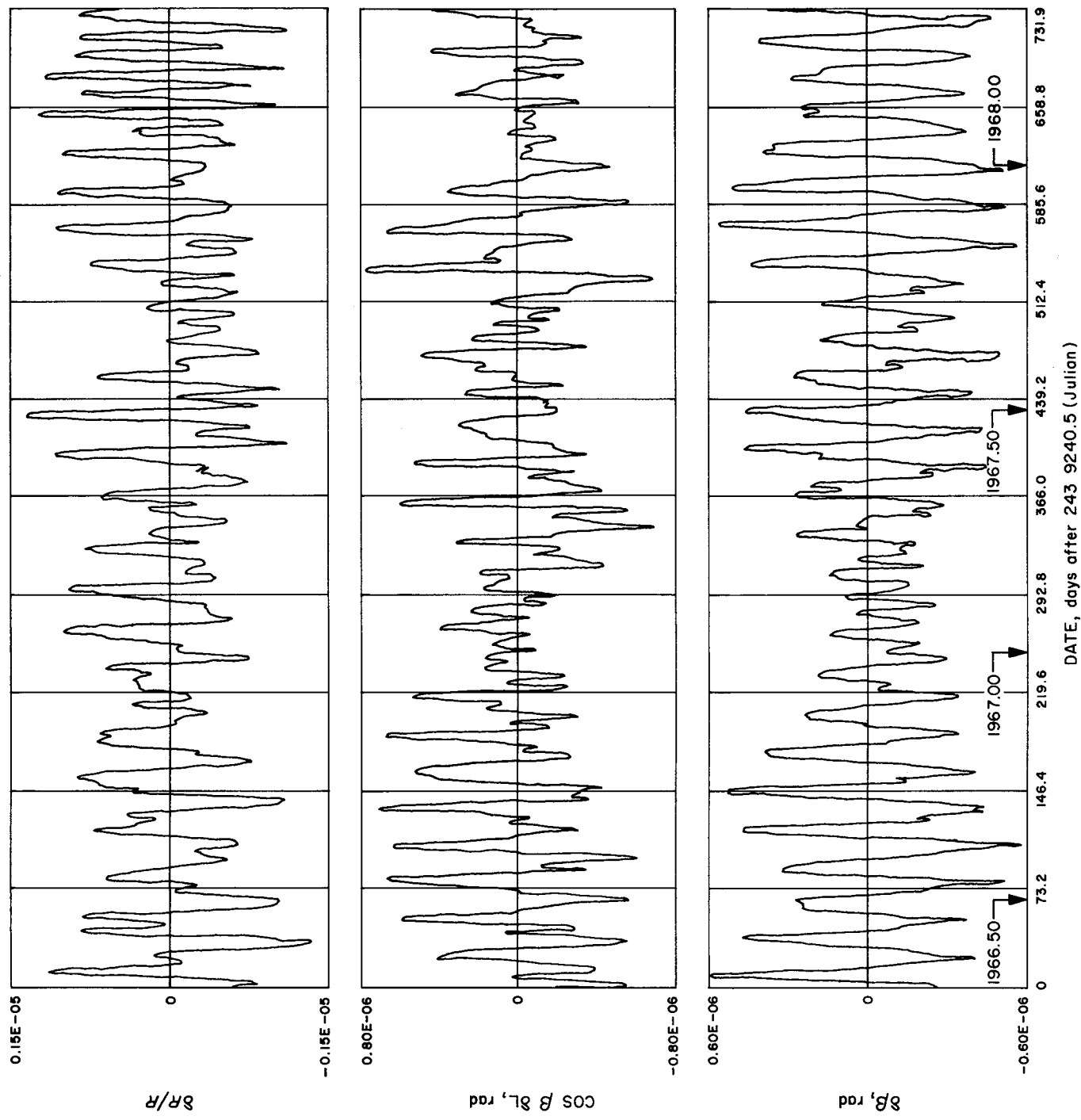


Fig. 2. Polar coordinate residuals

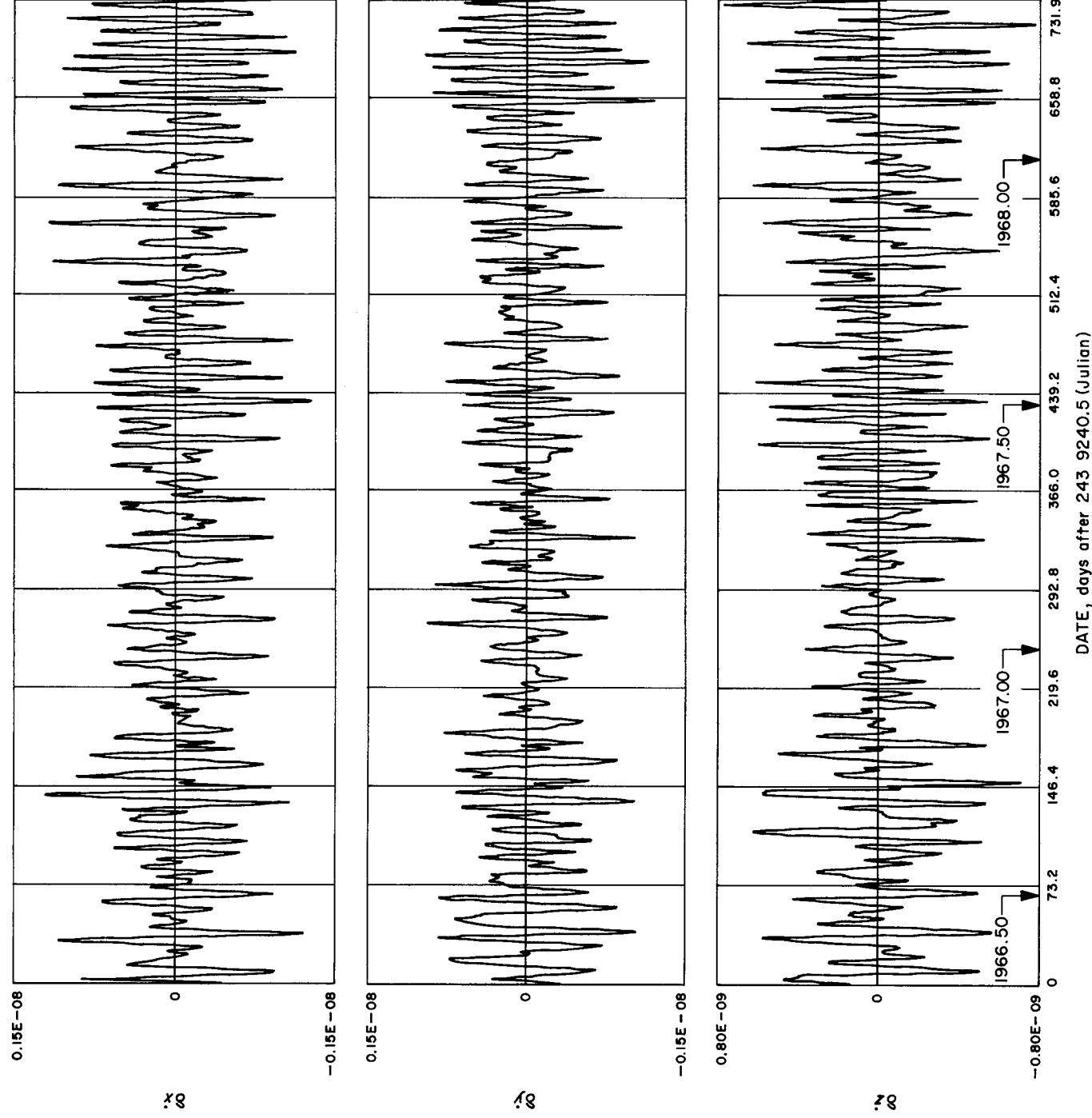


Fig. 3. Velocity residuals

C. Systems of Ordinary Differential Equations With an Algebraic Constraint, A. J. Semtner

Some recent work at JPL⁴ has dealt with the problem of determining whether a given system of ordinary differential equations

$$\left. \begin{aligned} \dot{x}_1 &= X_1(x_1, \dots, x_n) \\ &\vdots \\ \dot{x}_n &= X_n(x_1, \dots, x_n) \end{aligned} \right\} \quad (1)$$

has a solution which lies on a given surface

$$\Xi(x_1, \dots, x_n) = 0 \quad (2)$$

such a solution being called a *selected solution*. In the case where explicit solutions of Eq. (1) are available (including dependence on initial values), the problem can be resolved easily by a direct substitution into Eq. (2). In the more common case where solutions are not available, other methods have been developed which deal only with the functions X_1, \dots, X_n and Ξ . A method developed by E. M. Keberle constructs higher-order directional derivatives of Ξ with respect to the direction field X_1, \dots, X_n until a functional dependency is reached. The question of the existence of selected solutions can be answered, provided it is possible to invert these derivatives. A variation of this procedure will be discussed here in some detail. A theorem on which it is grounded will be proved and illustrated. Also, a situation in which this variation is not applicable will be discussed.

1. Functional Dependency of Directional Derivatives

The operator of directional differentiation with respect to the vector field X_1, \dots, X_n is defined by

$$L = X_1 \frac{\partial}{\partial x_1} + \dots + X_n \frac{\partial}{\partial x_n}$$

When L is applied repeatedly to the function Ξ , a sequence of scalar functions $\{L^i \Xi \mid i = 1, 2, \dots\}$ is generated. For short, we use the notation $\Xi^i = L^i \Xi$.

Suppose now that for the given function Ξ and vector field X_1, \dots, X_n , the following conditions hold in some region R of Euclidean n -space:

(i) The functions $\Xi, \Xi^1, \dots, \Xi^{t-1}$ are functionally independent in R , i.e., their Jacobian relative to some set of

variables is non-zero at each point of R . To simplify notation, assume that

$$\frac{\partial(\Xi, \dots, \Xi^{t-1})}{\partial(x_1, \dots, x_t)} \neq 0 \quad \text{in } R$$

(ii) The functions Ξ, \dots, Ξ^t are functionally dependent in the sense that

$$\frac{\partial(\Xi, \dots, \Xi^t)}{\partial(x_1, \dots, x_t, x_k)} = 0 \quad \text{in } R$$

for $k = t + 1, \dots, n$.

The conditions guarantee the existence of a function H defined on $S = (\Xi \times \dots \times \Xi^{t-1})(R)$ such that

$$H(\Xi(\vec{x}), \dots, \Xi^{t-1}(\vec{x})) = \Xi^t(\vec{x}) \quad (3)$$

for every $\vec{x} \in R$. In fact, H can be constructed as follows. Introduce new variables

$$\left. \begin{aligned} \xi_0 &= \Xi(x_1, \dots, x_n) \\ &\vdots \\ \xi_{t-1} &= \Xi^{t-1}(x_1, \dots, x_n) \end{aligned} \right\} \quad (4)$$

Because of (i), there is a neighborhood $N_{\vec{x}_0}$ of each point $\vec{x}_0 \in R$ in which the functions Eq. (4) can be inverted to obtain

$$\left. \begin{aligned} \hat{x}_1 &= x_1(\xi_0, \dots, \xi_{t-1}, x_{t+1}, \dots, x_n) \\ &\vdots \\ \hat{x}_t &= x_t(\xi_0, \dots, \xi_{t-1}, x_{t+1}, \dots, x_n) \end{aligned} \right\} \quad (5)$$

Now define a function H by

$$\begin{aligned} H(\xi_0, \dots, \xi_{t-1}, x_{t+1}, \dots, x_n) \\ = \Xi^t(\hat{x}_1(\xi_0, \dots, \xi_{t-1}, x_{t+1}, \dots, x_n), \dots, \\ \hat{x}_t(\xi_0, \dots, \xi_{t-1}, x_{t+1}, \dots, x_n), x_{t+1}, \dots, x_n) \end{aligned}$$

Since the functions in Eq. (5) are inverse to those in Eq. (4), it follows that

$$H(\Xi(x_1, \dots, x_n), \dots, \Xi^{t-1}(x_1, \dots, x_n), x_{t+1}, \dots, x_n) = \Xi^t(x_1, \dots, x_n) \quad (6)$$

This relation can be successively differentiated with respect to x_1, \dots, x_n to yield a system of linear equations for unknowns $\partial H / \partial \xi_0, \dots, \partial H / \partial \xi_{t-1}, \partial H / \partial x_{t+1}, \dots, \partial H / \partial x_n$. Because of (ii), we obtain by Cramer's rule that $\partial H / \partial x_{t+1} = 0, \dots, \partial H / \partial x_n = 0$.

⁴Technical Report 32-1101; SPS 37-28, Vol. IV, p. 15; SPS 37-42, Vol. IV, pp. 1, 2, pp. 15-17; SPS 37-43, Vol. IV, pp. 33-41.

Thus H can be regarded as a function only of ξ_0, \dots, ξ_{t-1} with domain $(\Xi \times \dots \times \Xi^{t-1})(N_{x_0})$. Equation (6) shows that H gives the desired functional dependency, Eq. (3). H can now be defined on all of S by piecing together H 's obtained from a covering of R by neighborhoods N_{x_0} . That this defines H in a consistent manner follows from the uniqueness of functional inversion. The above considerations on functional dependency apply, of course, to any functions satisfying (i) and (ii) and not just to a set of directional derivatives. They have been included to indicate explicitly how the function H can be constructed.

2. Existence of Selected Solutions

The following theorem connects the existence of selected solutions of Eq. (1) with properties of the functions Ξ, \dots, Ξ^{t-1} , and H :

Assuming that conditions (i) and (ii) are satisfied, there is a selected solution in R of Eqs. (1) if and only if

(iii) *There exists a point $(\bar{x}_1, \dots, \bar{x}_n) \in R$ such that*

$$\Xi(\bar{x}_1, \dots, \bar{x}_n) = \dots = \Xi^{t-1}(\bar{x}_1, \dots, \bar{x}_n) = 0$$

and

(iv) $H(0, \dots, 0) = 0$

Proof: Suppose a selected solution exists, namely

$$x_1(t), \dots, x_n(t)$$

where t runs through some interval I . The real-valued function

$$f(t) = \Xi[x_1(t), \dots, x_n(t)]$$

vanishes identically in I . All derivatives of f must also vanish. Since

$$\frac{d^k f}{dt^k}(t) = \Xi^k(x_1(t), \dots, x_n(t))$$

for $k = 0, 1, \dots$, conditions (iii) and (iv) are evidently satisfied.

Now suppose that the two conditions hold. By means of the change of variables in Eq. (4) at the point $(\bar{x}_1, \dots, \bar{x}_n)$,

the system Eq. (1) of ordinary differential equations is transformed to the equivalent system

$$\left. \begin{aligned} \dot{\xi}_0 &= \xi_1 \\ &\vdots \\ \dot{\xi}_{t-2} &= \xi_{t-1} \\ \dot{\xi}_{t-1} &= H(\xi_0, \dots, \xi_{t-1}) \\ \dot{x}_{t+1} &= \hat{X}_{t+1}(\xi_0, \dots, \xi_{t-1}, x_{t+1}, \dots, x_n) \\ &\vdots \\ \dot{x}_n &= \hat{X}_n(\xi_0, \dots, \xi_{t-1}, x_{t+1}, \dots, x_n) \end{aligned} \right\} \quad (7)$$

Here $\hat{X}_k(\xi_0, \dots, \xi_{t-1}, x_{t+1}, \dots, x_n) = X_k(\hat{x}_1(\xi_0, \dots, \xi_{t-1}, x_{t+1}, \dots, x_n), \dots, \hat{x}_t(\xi_0, \dots, \xi_{t-1}, x_{t+1}, \dots, x_n), x_{t+1}, \dots, x_n)$. The simple form of the first part of Eq. (7) is due to the fact that the ξ 's are directional derivatives of successively increasing order.

Because $H(0, \dots, 0) = 0$, a solution of the above is given by $\xi_0(t) = \dots = \xi_{t-1}(t) = 0$, supplemented by a solution $(x_{t+1}(t), \dots, x_n(t))$ of the reduced system

$$\left. \begin{aligned} \dot{x}_{t+1} &= \hat{X}_{t+1}(0, \dots, 0, x_{t+1}, \dots, x_n) \\ &\vdots \\ \dot{x}_n &= \hat{X}_n(0, \dots, 0, x_{t+1}, \dots, x_n) \end{aligned} \right\} \quad (8)$$

which starts at the point $(\bar{x}_{t+1}, \dots, \bar{x}_n)$.

A solution of Eq. (1) is then given by

$$\left. \begin{aligned} x_1(t) &= \hat{x}_1(0, \dots, 0, x_{t+1}(t), \dots, x_n(t)) \\ &\vdots \\ x_t(t) &= \hat{x}_t(0, \dots, 0, x_{t+1}(t), \dots, x_n(t)) \\ x_{t+1}(t) & \\ &\vdots \\ x_n(t) & \end{aligned} \right\}$$

This solution starts at $(\bar{x}_1, \dots, \bar{x}_n)$ and satisfies Eq. (2). It is thus a selected solution.

3. An Example

A simple example derived by J. S. Zmuidzinas⁵ will serve to illustrate the above theorem. (See also an example involving the Kepler problem in SPS 37-44, Vol. IV, pp. 1, 2.) We ask whether there is a solution of

$$\left. \begin{aligned} \dot{x}_1 &= x_1^2 + x_2^2 \\ \dot{x}_2 &= 2x_1x_2 \end{aligned} \right\}$$

⁵SPS 37-44, Vol. IV, pp. 265-269.

which satisfies the constraint

$$x_1 + x_2 = 0$$

In this case,

$$\Xi(x_1, x_2) = x_1 + x_2 \text{ and } \Xi^1(x_1, x_2) = x_1^2 + x_2^2 + 2x_1x_2$$

Conditions (i) and (ii) are satisfied when $\ell = 1$ and $R =$ the entire plane. The function H is clearly given by $H(t) = t^2$. Since $\Xi(\bar{x}_1, \bar{x}_2) = 0$ whenever (\bar{x}_1, \bar{x}_2) is a point on the line $x_1 + x_2 = 0$ and since $H(0) = 0$, a selected solution exists. In fact, a selected solution starting at the point $(c, -c)$ for $c > 0$ is given by

$$x_1(t) = \frac{1}{2t + 1/c}, \quad x_2(t) = \frac{-1}{2t + 1/c}$$

4. A Remark

Condition (iii) of the theorem follows from condition (iv) if we agree that H is defined only on

$$S = (\Xi x \cdots x \Xi^{l-1})(R)$$

It is explicitly stated to avoid a possible error when H has a natural form which can be evaluated at $(0, \cdots, 0)$, even though $(0, \cdots, 0)$ is not in S . For example, in the case

$$\left. \begin{array}{l} \dot{x}_1 = x_2^2 \\ \dot{x}_2 = 1 \end{array} \right\} \quad x_1 = 0 \quad (9)$$

we find that (i) and (ii) hold for $\ell = 1$ and

$$R = \{(x_1, x_2) | x_2 > 0\}$$

The function H is given by $H(t) = 2t^{1/2}$. Although $H(0) = 0$ in the extended sense, there is no selected solution in R because (iii) is violated.

5. The Case of Functional Degeneracy

The theorem does not cover the case of an "exceptional point" $(\bar{x}_1, \cdots, \bar{x}_n)$ at which Ξ, \cdots, Ξ^{l-1} are functionally independent and every

$$\frac{\partial(\Xi, \cdots, \Xi^l)}{\partial(x_1, \cdots, x_l, x_k)} = 0$$

at $(\bar{x}_1, \cdots, \bar{x}_n)$ but not all

$$\frac{\partial(\Xi, \cdots, \Xi^l)}{\partial(x_1, \cdots, x_l, x_k)}$$

are identically zero in a neighborhood of $(\bar{x}_1, \cdots, \bar{x}_n)$. In this case, no function H satisfying (3) can be found, yet the function chain cannot be carried further.

It is possible that there is a selected solution of Eq. (1) whose trajectory consists entirely of exceptional points. The existence of such a solution cannot be ascertained by using the theorem. For instance, such a selected solution for the situation

$$\left. \begin{array}{l} \dot{x}_1 = x_2 \\ \dot{x}_2 = x_1 \end{array} \right\} (x_1 - x_2)^2 = 0$$

is given by $x_1(t) = x_2(t) = e^t$. On the other hand, there may not be a selected solution consisting of exceptional points. Thus, in Eq. (9), any point of the form $(c, 0)$ is exceptional, but there are no solutions of Eq. (9) anywhere in the plane.

To understand better what happens near an exceptional point $(\bar{x}_1, \cdots, \bar{x}_n)$, let us transform Eq. (1) as before by introducing the new variables in Eq. (4). The following system, equivalent to Eq. (1), is obtained:

$$\left. \begin{array}{l} \dot{\xi}_0 = \xi_1 \\ \vdots \\ \dot{\xi}_{l-2} = \xi_{l-1} \\ \dot{\xi}_{l-1} = \hat{\Xi}^l(\xi_0, \cdots, \xi_{l-1}, x_{l+1}, \cdots, x_n) \\ \dot{x}_{l+1} = \hat{X}_{l+1}(\xi_0, \cdots, \xi_{l-1}, x_{l+1}, \cdots, x_n) \\ \vdots \\ \dot{x}_n = \hat{X}_n(\xi_0, \cdots, \xi_{l-1}, x_{l+1}, \cdots, x_n) \end{array} \right\} \quad (10)$$

This system of equations differs from the system Eq. (7) in that the expression for $\dot{\xi}_{l-1}$ now depends on x_{l+1}, \cdots, x_n . It is no longer possible to view the transformed system as two "uncoupled" systems for the variables $(\xi_0, \cdots, \xi_{l-1})$ and (x_{l+1}, \cdots, x_n) , respectively. Previously, we could set $\xi_0 = \cdots = \xi_{l-1} = 0$ and solve the unconstrained system Eq. (8) for x_{l+1}, \cdots, x_n . Now we must solve Eq. (8) subject to the constraint

$$\hat{\Xi}^l(0, \cdots, 0, x_{l+1}, \cdots, x_n) = 0 \quad (11)$$

The problem now arises that $\partial \hat{\Xi}' / \partial x_k = 0$ at $(0, \dots, 0, \bar{x}_{t+1}, \dots, \bar{x}_n)$. This follows from the fact that each $[\partial(\Xi, \dots, \Xi') / \partial(x_1, \dots, x_t, x_k)] = 0$ at $(\bar{x}_1, \dots, \bar{x}_n)$. We cannot approach the new system (Eq. 8) with constraint (Eq. 11) by constructing a new chain of directional derivatives, since it is impossible to introduce even the constraint itself as a new variable near $(\bar{x}_{t+1}, \dots, \bar{x}_n)$. Until a method is devised to deal with this situation, the existence of selected solutions consisting of exceptional points will remain in doubt.

D. Reduction of a Problem in Relativistic Cosmology Using a Transformation Algorithm, A. J. Semtner

A system of nine ordinary differential equations has been derived by Estabrook, Wahlquist, and Behr (Ref. 1) to describe a cosmology with incoherent matter. The question has arisen whether there is a solution of these equations which satisfies four given algebraic constraints. This problem is shown in the following section 1 to break up

into a number of cases involving fewer differential equations and constraints. Each case can be treated using a transformation algorithm with a shortcut (Ref. 2 and SPS 37-43, Vol. IV, pp. 33-41) which reduces the number of differential equations at each step. The object of the transformation algorithm is to produce finally a system of equations in "partially separated" form. A simple test consisting of the application of the theorem of partial stationarity can then determine whether there is a selected solution of the original differential equations and the constraints in the given case. However, the present problem is of such algebraic complexity that hand calculations are inadequate to bring about the partially separated form. Still, the problem can be reduced considerably in each case before a situation is reached where machine methods must be used. This reduction is carried out in section 2 of this article for a typical case.

1. Original Equations Divided Into Cases

The system of differential equations is as follows:

$$\begin{aligned} \dot{\phi} &= \frac{-2\tau^2(c+a)}{c-a} + \frac{2v^2(a+b)}{a-b} - (\phi + \chi + \psi)\phi \\ &+ (-a+b+c)a + \frac{1}{4}\rho^{-1}\sigma^2(b-c)^2 - \frac{1}{4}\rho^{-1}\tau^2(c-a)^2 \\ &- \frac{1}{4}\rho^{-1}v^2(a-b)^2 + \frac{1}{2}\Lambda + \frac{1}{2}(\phi\chi + \chi\psi + \psi\phi - \sigma^2 - \tau^2 - v^2) \\ &+ \frac{3}{8}(a^2 + b^2 + c^2 - 2ab - 2bc - 2ca) \\ \dot{\sigma} &= \frac{-2v\tau(bc-a^2)}{(a-b)(c-a)} - \frac{\sigma(\chi-\psi)(b+c)}{(b-c)} + \frac{1}{2}\rho^{-1}v\tau(c-a)(a-b) - (\phi + \chi + \psi)\sigma \\ \dot{a} &= (\phi - \chi - \psi)a \end{aligned}$$

plus six more equations obtained from the above by simultaneous cyclic permutation of the triplets $(\phi\chi\psi)$, (abc) , $(\sigma\tau v)$. Λ is a constant. ρ is given by

$$4\rho = 2(\phi\chi + \chi\psi + \psi\phi - \sigma^2 - \tau^2 - v^2) - \frac{1}{2}(a^2 + b^2 + c^2) + ab + bc + ca - 2\Lambda$$

The constraints are

$$a(b-c)\sigma = 0$$

$$b(c-a)\tau = 0$$

$$c(a-b)v = 0$$

$$4\rho^2(\phi + \chi + \psi) - (b-c)^2\sigma^2\phi - (c-a)^2\tau^2\chi - (a-b)^2v^2\psi - 2(ab + bc + ca - a^2 - b^2 - c^2)\sigma\tau v = 0$$

Since each of the first 3 constraints is a product of 3 terms, there are 27 ways in which these constraints can be satisfied. Many of the ways give the same information or are equivalent after cyclic permutation. When these circumstances are taken into consideration, only 10 cases need to be treated:

I	$a = b = c = 0$	VI	$a = \tau = v = 0$
II	$a = b = 0$	VII	$a = b = c$
III	$b = c = \sigma = 0$	VIII	$a = b = c, v = 0$
IV	$a = c = v = 0$	IX	$b = c, \tau = v = 0$
V	$a = b = \tau = 0$	X	$\sigma = \tau = v = 0$

There will be a solution of the 9 differential equations and 4 constraints if, and only if, there is a solution of the 9 differential equations and the last constraint which satisfies the conditions of one of the 10 cases.

2. The Algorithm Applied to a Specific Case

Associated with each of the above cases is a smaller and simpler system of differential equations and constraints obtained by imposing the requirements of the given case on the nine differential equations and the fourth constraint. Some of the old differential equations will drop out or appear as new constraints when the substitutions are made. For example, in case X, the equations for $\dot{\sigma}$, $\dot{\tau}$, and \dot{v} drop out, since these equations reduce to the form $0 = 0$ when we put σ , τ , v , $\dot{\sigma}$, $\dot{\tau}$, \dot{v} all zero. On the other hand, in case III, the old equation for $\dot{\sigma}$ becomes the constraint $0 = -2v\tau - \frac{1}{2}\rho^{-1}v\tau$ when we put b , c , σ , \dot{b} , \dot{c} , $\dot{\sigma}$ all zero.

Let us now apply the transformation algorithm to case X. Here, we are dealing with the differential equations:

$$\begin{aligned}\dot{\phi} &= -(\phi + \chi + \psi)\phi + (-a + b + c)a + \frac{1}{2}\Lambda \\ &+ \frac{1}{2}(\phi\chi + \chi\psi + \psi\phi) \\ &+ \frac{3}{8}(a^2 + b^2 + c^2 - 2ab - 2bc - 2ca) \\ \dot{a} &= (\phi - \chi - \psi)a\end{aligned}\quad (1)$$

plus four more equations from permutations. The constraint is that $4\rho^2(\phi + \chi + \psi) = 0$. Since the situation $\rho = 0$

is not of physical importance, ρ being the mass density, we replace the above with

$$\phi + \chi + \psi = 0 \quad (2)$$

At each step of the transformation algorithm the differential equations are transformed by introducing the constraints as new variables. Then the constraint variables and their derivatives are set zero to give a new constrained system, which has a solution if, and only if, the previous system does. One step of the algorithm applied to Eqs. (1) and (2) to eliminate ϕ gives

$$\begin{aligned}\dot{\chi} &= (-b + c + a)b - \frac{1}{2}(\chi^2 + \psi^2 + \chi\psi - \Lambda) \\ &+ \frac{3}{8}(a^2 + b^2 + c^2 - 2ab - 2bc - 2ca) \\ \dot{\psi} &= (-c + a + b)c - \frac{1}{2}(\chi^2 + \psi^2 + \chi\psi - \Lambda) \\ &+ \frac{3}{8}(a^2 + b^2 + c^2 - 2ab - 2bc - 2ca)\end{aligned}\quad (3)$$

$$\dot{a} = -2(\chi + \psi)a$$

$$\dot{b} = 2\chi b$$

$$\dot{c} = 2\psi c$$

with constraint

$$\frac{1}{8}(a^2 + b^2 + c^2 - 2ab - 2bc - 2ca) - \frac{3}{2}(\chi^2 + \psi^2 + \chi\psi - \Lambda) = 0 \quad (4)$$

A second step of the algorithm applied to Eqs. (3) and (4) yields

$$\begin{aligned}\dot{\chi} &= 2c \pm \frac{1}{8}[bc + 3(\chi, \psi)]^{1/2} - \frac{1}{2}(\chi, \psi) \\ &+ \frac{3}{512}[bc + 3(\chi, \psi)]^{1/2} \\ \dot{\psi} &= -2b \pm \frac{1}{8}[bc + 3(\chi, \psi)]^{1/2} - \frac{1}{2}(\chi, \psi) \\ &+ \frac{3}{512}[bc + 3(\chi, \psi)]^{1/2} \\ \dot{b} &= 2\chi b \\ \dot{c} &= 2\psi c\end{aligned}\quad (5)$$

with constraint

$$\begin{aligned}
 & 7(\chi b^2 + \psi c^2) - \frac{8227}{512}(\chi + \psi)bc \\
 & \pm \left[\frac{35}{4}(-\chi + \psi) + \frac{1}{8}[bc - 3(\chi, \psi)]^{\frac{1}{2}} \left(\frac{3}{4}\chi - \frac{19}{8}\psi \right) \right] b \\
 & \pm \left[\frac{35}{4}(-\chi + \psi) + \frac{1}{8}[bc - 3(\chi, \psi)]^{\frac{1}{2}} \left(-\frac{19}{4}\chi + \frac{3}{4}\psi \right) \right] c \\
 & + \frac{2199}{512}(\chi, \psi)(\chi + \psi) = 0 \quad (6)
 \end{aligned}$$

Here $(\chi, \psi) = \chi^2 + \psi^2 + \chi\psi - \Lambda$. The \pm sign indicates a division into two subcases due to the quadratic nature of constraint, Eq. (4).

The explicit reduction of case X by the transformation algorithm is not continued further, since constraint, Eq. (6), cannot readily be inverted with respect to any of its variables for use in transforming Eq. (5). Machine methods will be needed for the termination of the algorithm.

References

1. Estabrook, F. B., Wahlquist, H. D., and Behr, C. G., "Dyadic Analysis of Spatially Homogeneous World Models," *J. Math. Phys.* (in press).
2. Keberle, E. M., *Partial Stationarity as a Compatibility Criterion for Overdetermined Systems of Ordinary Differential Equations, Related to Non-Differential Constraints*, Technical Report 32-1101. Jet Propulsion Laboratory, Pasadena, Calif., Mar. 30, 1967.

IV. System Design and Integration

PROJECT ENGINEERING DIVISION

A. Advanced Development at the System Level,

K. Casani and J. Gardner

1. Introduction

Historically, advanced development at JPL has been conducted at the subsystem level in order to place the organization in the best position to undertake a project at an appropriate time. This approach tends to advance most disciplines on a broad front, sometimes unnecessarily and sometimes leaving a key area undeveloped. As a means of protecting against these dangers, focusing the JPL advanced development program, and gaining some insight into system-level problems, a new concept was selected: system-level advanced development.

The first task selected was to pursue the system design of a planetary entry capsule. The Mars 1971 opportunity was used to give some real constraints to the system design. This work, presently being carried out under the CSAD program, has the following objectives:

- (1) To provide a means for gaining experience in several critical and new technologies related to planetary capsule missions.

- (2) To develop an understanding of the subsystems such that realistic performance estimates can be made.
- (3) To obtain an improved understanding of planetary entry capsule system design and integration problems.

2. Description and Status of the Entry and Lander Systems

The capsule¹ is composed of an entry capsule and a lander. The entry capsule must be capable of surviving the entry, aerodynamic braking, and heating. The lander, carried through atmospheric entry by the entry capsule, must be capable of surviving landing impact after a parachute descent.

The initial program was to design, integrate, develop, and fabricate an entry capsule with at least the following subsystems: aeroshell, relay communication, power, and mass spectrometer. This entry capsule was to be subjected to a series of functional tests, a sterilization heat cycle, and selected environmental tests to demonstrate the soundness

¹Described in "Mars '71 Technical Study," Aug. 15, 1966, and "Mars '71 Technical Study, Addendum 1," Dec. 12, 1966 (JPL internal publications).

of the design and point out system and subsystem incompatibilities.

The initial program for the lander was to design, integrate, develop, and fabricate a lander capable of being integrated into the entry capsule and withstanding the environmental and impact constraints placed on it during flight, entry, parachute descent, and landing. The lander was initially planned to have at least the following subsystems: impact limiter, direct radio, power, sequencer and timer, orientation scheme, and simulated parachute and mortar. Functional tests, a sterilization heat cycle, and selected environmental tests were planned, as well as an impact test as an additional test requirement.

Significant progress on the entry capsule and lander has been made:

- (1) The capsule system design is well into its final stages after undergoing many design changes and reviews.
- (2) The capsule system functional block diagram has undergone several iterations due to changes from the conceptual design and now reflects a sound system design. The interdependence of the subsystems in terms of power and command capabilities is indicated.
- (3) A first-level capsule system logic diagram has been initiated. This diagram, complementary to the capsule system functional block diagram, is useful to identify weak points in the design.
- (4) A state-time diagram has been initiated to identify the sequence of events and the state of each function in every subsystem at any time.
- (5) A failure-analysis time diagram, showing the sequence of events and the temporal relationship between primary and backup commands, has been started.
- (6) A full-scale mockup of the capsule system has been fabricated. This mockup has been useful in identifying the class and order of assembly and operational problems to be encountered. It will also be used for the actual system cabling layout.
- (7) A digital computer program has been developed to simulate landing. This program calculates the probability of landing success for a lander designed to a specific set of design constraints. This is done by statistically defining the landing conditions from

probability density functions of the atmospheric surface temperatures, pressures, molecular weights, winds, and surface slopes. These analytical results have been compared with actual hardware test results, and indications are that an adequate margin exists in the lander design.

The CSAD program² is approaching the period of hardware delivery, the finalization of operational-support-equipment requirements, and the development of detailed test plans and schedules.

3. Description of the FM

Construction of an FM has been initiated. The FM is a functioning engineering model patterned after the capsule system described in the JPL internal publications cited in Footnote 1. The mission described in that document, considered representative of planetary capsule missions, was used, as required, to generate a set of representative mission requirements needed to implement a meaningful system design. It was not intended that the FM be a complete working capsule system, but that it demonstrate specific subsystem capabilities unique to a planetary capsule mission and not as well understood as those of interplanetary spacecraft. The objectives to be met by the FM were:

- (1) To demonstrate required key subsystem technologies in a functional capsule system.
- (2) To expose a functional capsule system to the terminal sterilization environment.
- (3) To gain experience in the system test and operations of a planetary capsule.

Since it was not practical, with the resources available, to design, fabricate, and test a complete engineering-prototype capsule system, only those elements that would contribute directly to accomplishing the stated program objectives were intended to be incorporated in the FM. The capability of the FM varies from element to element in that some subsystems are non-functional mockups, while others are flight-designed although not flight-qualified. The fully functional subsystems are as follows:

- (1) Entry structure.
- (2) Entry radio subsystem.

²Initial schedule defined in "Capsule System Advanced Development Program Guidelines," Apr. 25, 1967 (JPL internal publication).

- (3) Entry battery.
- (4) Entry power control unit.
- (5) Entry cabling.
- (6) Entry data subsystem (breadboard, external).
- (7) Lander structure.
- (8) Lander radio.
- (9) Lander battery.
- (10) Lander power control unit.
- (11) Lander sequencer and timer.
- (12) Lander cabling.
- (13) Aeroshell.
- (14) Sterilization canister.
- (15) Impact limiter.
- (16) Mass spectrometer (some external components).
- (17) Instrument boom.
- (18) Landing and orientation sensors.
- (19) Maneuver pyrotechnics.

This listing reflects a substantial increase in the number of functional subsystems over the initial program projections. Subsystems supporting the FM in a partially functional capacity are the entry temperature control, lander temperature control, and lander wind instrument subsystems. The parachute initiator, entry sequencer and timer,

entry sensor, and entry timer subsystems are simulated by operational support equipment. The non-functional subsystems (mockups) are the following:

- (1) Propulsion.
- (2) Separation-initiated timers.
- (3) Spin-despin.
- (4) Radiometer.
- (5) Entry water vapor instrument.
- (6) Lander temperature sensor.
- (7) Lander water vapor instrument.
- (8) Lander data subsystem.
- (9) Parachute subsystem.
- (10) Gas chromatograph.
- (11) Entry aerometry experiment.
- (12) Lander pressure probe.

An introductory discussion of the CSAD program was presented here. The changes in the entry capsule system design from the conceptual design,³ the discovery of new system-level problems and solutions, and the effect of the test program through subsystem and system evaluation will be described in future issues of the SPS, Vol. III.

³Outlined in "Mars '71 Technical Study," Aug. 15, 1966, and "Mars '71 Technical Study, Addendum 1," Dec. 12, 1966 (JPL internal publications).

PRECEDING PAGE BLANK NOT FILMED.

V. Spacecraft Power

GUIDANCE AND CONTROL DIVISION

A. Sterilizable Battery, R. Lutwack

1. Development of Separators for Sterilizable Batteries

The research and development program for a separator for the heat sterilizable Ag-Zn battery comprises contracts with Monsanto Research Corp., Westinghouse Electric Corp., the Southwest Research Institute, and Narmco Division, Whittaker Corp. In addition, separator evaluations are being done at ESB, Inc., and a research and evaluation program is being conducted at JPL.

a. JPL Contract 951524. In this contract with Monsanto Research Corp., ligand-containing polymers are being investigated. Films have been prepared from styrene/maleic anhydride, styrene/maleic anhydride/methyl methacrylate, and 2-vinylpyridine/methyl methacrylate systems. The electrical properties of the terpolymers were improved by incorporating methyl acrylate, but increased solubility in the KOH solution also occurred; 2-chloroethylvinyl ether is being used as a crosslinking agent in efforts to reduce solubility. The flexibility of films fabricated from styrene/maleic anhydride has been increased by using preparations of higher molecular weight polymers in forming the films.

b. JPL Contract 951525. In this contract with Westinghouse Electric Corp., the techniques for the preparation of membranes composed of a matrix of Webril, a polysulfone binder, and a zirconium oxide filler are being investigated. A laboratory semicontinuous coating apparatus is being used to study the effects of varying the extractant solution, drying methods, mixture compositions, filler loading on film thickness, KOH diffusion time, and resistivity. These are the tentative conclusions: (1) the diffusion times are faster and there are no induction times for air-dried films; (2) air-dried films have lower resistance values than extracted films; (3) diffusion tests show that films can be made which are free of major defects; (4) fast diffusion times are concomitant with low resistance values.

c. JPL Contract 951966. This is a new contract with the Monsanto Research Corp. for the development of a polyethylene-acrylic acid-type separator material without the use of radiation. A free radical high pressure polymerization of ethylene-methyl acrylate copolymers, which are then hydrolyzed to the potassium salt, will be used. Crosslinking of the copolymer will be done through the use of peroxides and heat. If required, materials can be strengthened by blending with ethylene-vinyl acetate

copolymers or carbon black. In this program, Monsanto will determine the best copolymer composition, the need for reinforcing resins and fillers, the optimum crosslinking agent concentration, and the best hydrolysis parameters. The physical and chemical stabilities and the electrical resistance of the separator will be measured.

d. JPL Contract 951718. In this contract with the Southwest Research Institute (SwRI), the first phase was a detailed study of the basic parameters used in the grafting step of polyethylene-based, radiation-grafted and crosslinked battery separator material. These additional studies are under way: (1) the optimum crosslinking and grafting solution compositions will be determined; (2) standards for the various wash and rinse solutions will be established; (3) electron-beam crosslinking of the film will be attempted; and (4) the usefulness of substituted acrylic acids for grafting will be studied. The scale-up of the basic procedure is also being developed. SwRI will supply JPL with limited quantities of the separator material.

e. JPL Contract 951091. This is a contract with the Narmco Division of the Whittaker Corp. After a scale-up phase in the development of poly 2,2'-hexamethylene 5,5' bis (1- β carboxyethyl) benzimidazole as a separator material, sample quantities were delivered to JPL and are being tested for compatibility with the Ag-Zn system.

Work on 2,2'-octamethylene 5,5'-bibenzimidazole as a potential case material is also progressing. Studies indicate that solvent bonding with a solution of formic acid in dimethylformamide, followed by a heat curing cycle, produces bonds with a shear strength of 1300 psi. A study of injection molding for this polymer is now under way.

f. Separator evaluation program at ESB, Inc. The separator material prepared by Southwest Research Institute from polyethylene film by grafting with acrylic acid and crosslinking with divinylbenzene, using Co⁶⁰ irradiation for both processes, is being used exclusively in all of the Ag-Zn cell design and development. The evaluation of this material is thus concomitant with the evaluation of cell designs.

g. Separator program at JPL. Materials which show promise as heat sterilizable battery separators are being tested continuously at JPL. Screening tests include resistance measurements, tensile strength determinations, sterilization, and in-cell testing. The in-cell testing is further subdivided into cycle tests and stand tests of cells. Silver migration measurements will be added to the tests as soon

as the required equipment arrives. Spectrophotometric methods are being investigated in the search for a non-destructive test method. These tests are also used to provide quality assurance for production separator materials.

2. Research and Development of the Sterilizable Battery

a. Ag-Zn and Ag-Cd batteries (JPL Contract 951296 with ESB, Inc.). This is a research and development program for sealed Ag-Zn and Ag-Cd cells, which provide satisfactory electrical performance after heat sterilization. The first phase of the program, comprising tasks for electrochemistry, cell cases and sealing, and cell fabrication and testing, is nearly completed. These conclusions have been obtained as follows: (1) Standard ESB Ag electrodes are satisfactory after heat sterilization, but additional supports are necessary to meet high shock requirements; (2) HgO, which is in the usual Zn electrode to prevent H₂ evolution, cannot be a constituent of heat sterilizable Zn electrodes; HgO is sufficiently soluble in KOH solution at the sterilization temperature to interact with the Ag plate and decrease its capacity; (3) a substitute for HgO, which does not have this characteristic, has been found; (4) Zn electrodes also require additional supports to meet high shock requirements; (5) polyethylene separators survive sterilization and are being used in the Ag-Zn development cells; (6) Kendall Mills' EM 476 is a satisfactory absorber; (7) cases molded from polyphenylene oxide and sealed with epoxy resin can be sterilized, decontaminated, and shocked to 2500 g without rupture; (8) a modified charging procedure must be used to prevent H₂ gas evolution.

The second phase of the ESB program is for the design, fabrication, and testing of batteries, which in general are grouped in the classifications of primary-shock resistant, primary, primary-high rate, and secondary batteries. This phase is under way, utilizing the information obtained in Phase I.

b. Ni-Cd battery (JPL Contract 951972 with Texas Instruments, Inc.). This is a research and development program for heat sterilizable Ni-Cd cells. Tasks for separator evaluation, development of cases and seals, and characterization of Ni and Cd plates are under way.

B. Solar Cell Standardization, R. F. Greenwood

1. Introduction

Beginning as a method to more accurately predict the output power of a solar array in space, the solar cell

standardization program has expanded to include the calibration of solar cells for use in setting up artificial light sources and the testing and calibration of newly developed solar cells. High-altitude balloons which ascend above 97% of the earth's atmosphere are used to accomplish this task.

2. Results of 1967 Balloon Flights

During July and August 1967, a series of four balloon flights was successfully completed. Solar cells calibrated on these flights were supplied by the NASA Ames Research Center, The NASA Langley Research Center, Johns Hopkins University, the Air Force Aero Propulsion Laboratory, and JPL. The cells supplied represented a fairly large cross-section of solar cell types. Data received were reduced to a usable form by a computer program. Analysis indicated that the data were generally of excellent quality, and good correlation resulted among the flights this year as well as between this series and previous series of flights.

Three of the flights were designed to reach 80,000 ft while one flight was designed to attain an altitude of 120,000 ft. Cells flown at 80,000 ft were reflown at the higher altitude to determine if any attenuation of sunlight was evident because of air mass or ozone layers between the two altitudes. A comparison of the data revealed no significant change of the short circuit current of the solar cells at 120,000 ft.

Measurements of short circuit current and open circuit voltage were made on special three-cell modules termed Isc-Voc transducers. Four of these transducers were calibrated at an 80,000-ft altitude and are intended to be used with the *Mariner* Mars 1969 flight program. The transducers are to be mounted on flight panels and will serve to evaluate the performance of the solar panels both in ground tests and in flight.

Plans are being formulated for a new series of flights during the summer of 1968. It is anticipated that other interested agencies will again participate in a cooperative effort and exchange of information program.

C. Feasibility Study: 30-W/lb Roll-up Solar Array, W. A. Hasbach

1. Introduction

The ever-increasing power requirement of the spacecraft, coupled with the launch vehicle stowage capability,

dictated the need to evaluate new techniques in the construction of solar power panels. Assuming that the solar cell itself has reached the maximum conversion efficiency for the manufacturing processes as we know them today, emphasis was placed upon the mechanical-structural aspects of the solar array. Large area lightweight panels will require special designs to survive the rigors of spacecraft launch and to meet the packaging and deployment concepts required to fit these arrays into the volume limitation imposed by the shroud-spacecraft interfaces. A program was initiated at JPL in June 1967, to investigate the feasibility of developing a 10-kW solar array which would have a specific power capability of 30 W/lb and be deployed after launch through a roll-out technique similar to the packaging employed in a window shade.

The objectives are: (1) 30 W/lb at a sun-probe distance of 1 AU based on an energy conversion of 10 W/ft², AMO at 55°C equivalent solar intensity (0.83 lb/ft²); (2) 250 ft² of array surface.

2. Approach

A request for technical proposals was solicited from several possible contractors. From the seven who responded, three were selected to perform a 12-mo feasibility study on roll-up arrays. The three contractors are Ryan Aeronautical Co., General Electric Co., and Fairchild-Hiller. At the conclusion of the study program each contractor will submit a scale model of his chosen design, which will demonstrate the mechanical performance features of the most promising design evaluated.

Ryan Aeronautical Co. is pursuing a two-boom deployment system, as shown in Fig. 1. This design was basically employed in the development of a small 50 ft² roll-out solar array developed under JPL contract 951107. This array uses two compressible beams which collapse about the drum as the array is retracted (Ref. 3, Table 1).

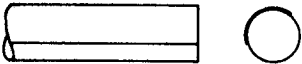
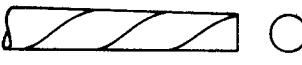
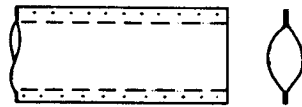
During the study phase of this contract the collapsible beam technology gained in the 50 ft² array will be expanded and improved both in the materials and structural aspect. The mechanical details of the proposed design are shown in Figs. 1 and 2. The deployment beam cross-section chosen is 1.7 in., fabricated of 0.003-in.-thick titanium (6 AL4V).

The wrap drum will be 12.00 in. in diameter, fabricated of magnesium alloy sheet perforated with lightning holes. Beryllium is also being evaluated. The end bulkheads are of aluminum honeycomb. The drum is mounted to the

support structure channels by a sliding fitting at each end. A spring is attached to each fitting and anchored at its other end to the structure. The spring load pulls the

drum tightly against the drive roller and idler. During retraction the beams build up on the drum and the drum center rises in the slides against the spring load.

Table 1. Extendible boom configurations

Ref. No.	Configuration	Illustration	Manufacturer	Description	Flight experience
1	Storage Tubular Extendible Member (STEM)		a. De Havilland Aircraft Canada (BeCu) b. GE Spacecraft Dept. (Molybdenum) c. Douglas Missile & Space Systems Div. (Fiberglass-Epoxy)	Preformed spring sheet stock unrolled from drum; overlapped into tube upon extension; can be perforated. Sometimes nested.	Many vehicles. Being used for ATS, RAE, GGTS, etc.
2	Spiral wrap		Hunter Spring Corp. Hatfield, Penn.	Stainless steel tape formed in spiral configuration and stored on drum. Reverts to spiral tube when extended.	None to date (development models built)
3	Flattened tube boom		a. Ryan Aircraft Corp. (welded edges)	Preformed metal tapes joined by welding at edges; flattened and rolled as two sheets on a drum. Expands when extended.	None to date (development models built)

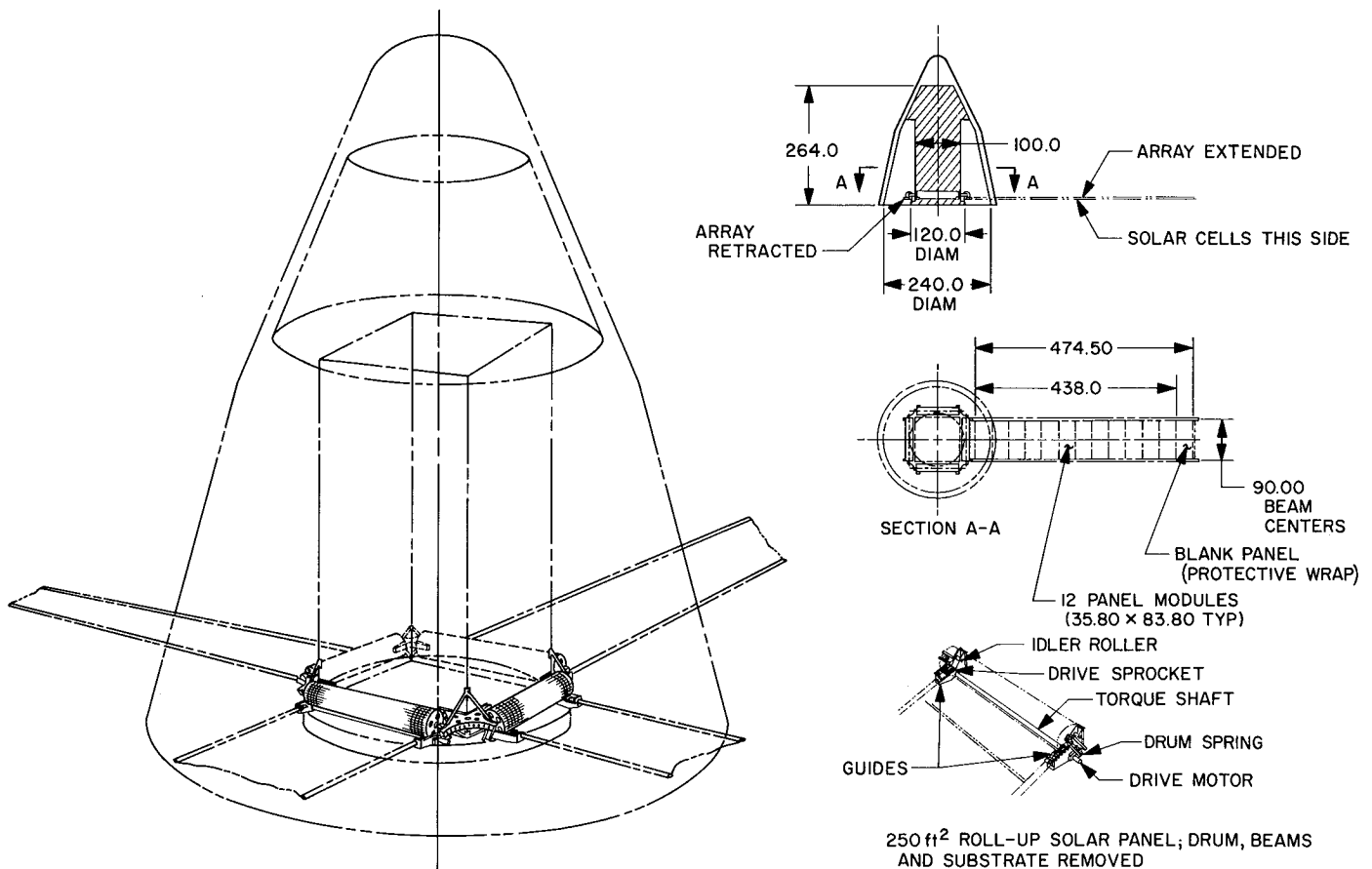
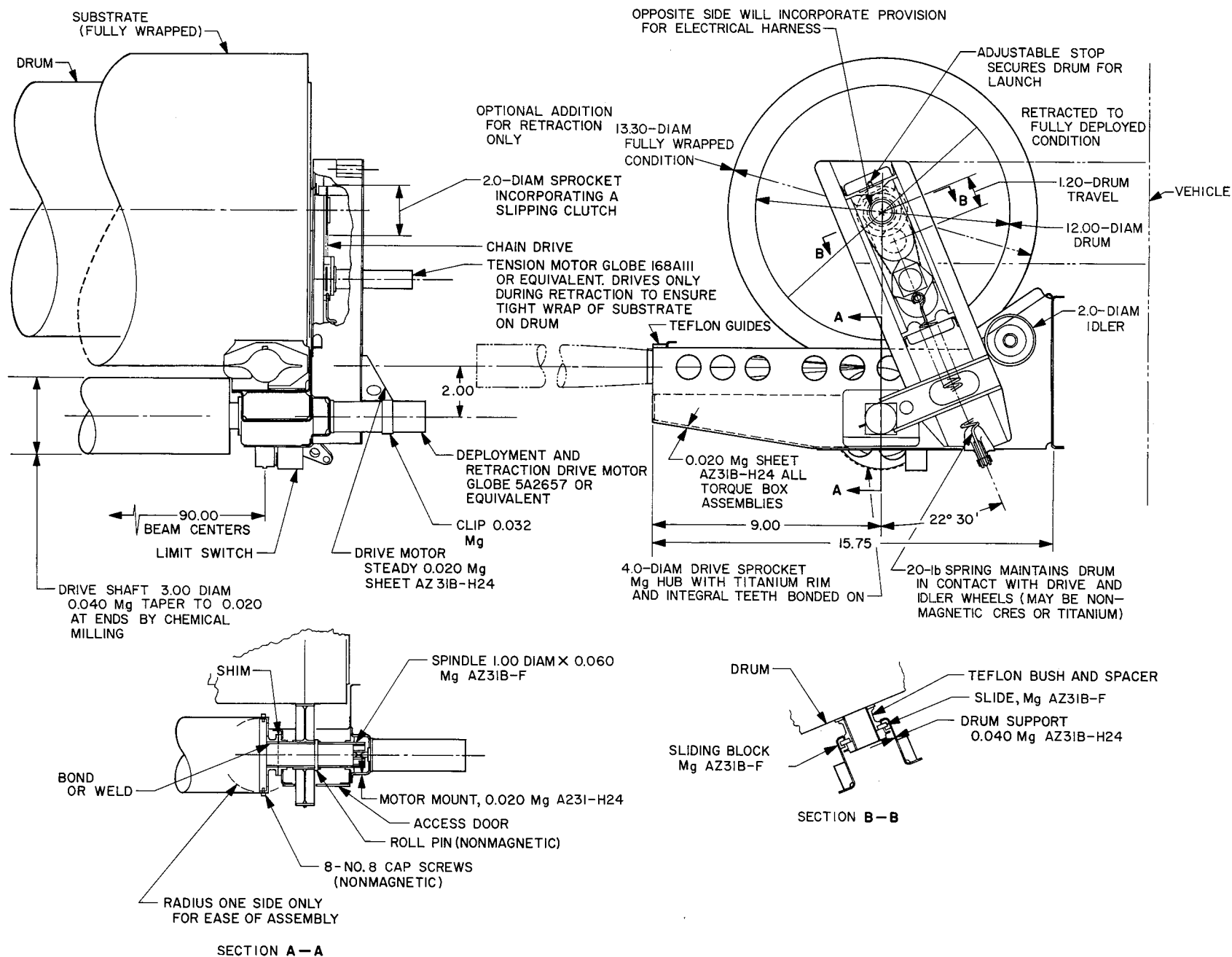


Fig. 1. Roll-up solar array installation (1000 ft²)

Fig. 2. Support structure and deployment mechanism (250 ft² roll-up solar panel)

The drive roller compresses the beam against the wrap drum and provides the tangent point from which the beam starts to expand to its open section. The drive roller incorporates a gear tooth drive, the teeth of which engage corresponding teeth in a rack gear welded along the beam. This provides a positive drive for the beam in both the deployment and retraction modes. Attached to the outer end of the left-hand sprocket is a motor with a planetary gear system to provide the necessary power drive. Attached to the inner side of the sprocket is a magnesium torque shaft which carries the drive to the opposite side beam, thus synchronizing both beams. The beam guides are Teflon machined blocks, two provided for each beam and shaped to conform to the beam profile.

Reversing the drive motor will retract the array; however, with this system there is a possibility of slip between the beam wraps which could cause slackness in the substrate wrap-up. To guard against this, an additional drive and slip clutch will drive the drum at a slightly faster rate than the beams are being driven during the retraction sequence. During deployment the clutch provides a brake on the drum to stop any unwinding action.

Ryan has stated that the choice of 0.0015-0.002 Kapton (H Film) for the substrate would prove satisfactory; however, additional investigations will continue. Various methods of substrate attachment to the beams are under investigation. At the present time, single clips spaced at 4.0- to 6.0-in. intervals are indicated. The Kapton substrate would be reinforced in the area of clip attachment. The final configuration for the electrical feedout has not been determined at this time. Two approaches are under

consideration: a spirally wound harness, and slip rings. Each is designed and in the process of evaluation by the reliability section.

Fairchild-Hiller is evaluating two solar array concepts, as shown in Figs. 3 and 4. Other than the extension mechanism, both deployment systems are similar. This sameness is illustrated in Figs. 3 and 4.

The array (250 ft²) is divided into four modularized subpanels. Once released, the array deploys outwardly, guided and supported by the extension mechanism. The rate of deployment will be governed by a centrifugal brake attached to the substrate roller. Rate control is needed on this design so that billowing of the substrate will not occur, and to provide a constant radial velocity of the extension arms. Once the array is deployed it is tensioned positively by a spring acting on the roller to absorb variations in length between solar panel substrate and extension arm frequency. Prior to deployment, the substrate on the roller, and the extension mechanisms, are both secured within the structure by the release mechanism which is designed to provide support for the load in all three axes. It will be operated by a pyrotechnic device or solenoid.

The contractor's present weight analysis of the array, excluding any mechanism, and based on today's technology, is 0.262 lb/ft². In 1 yr it will be 0.227 lb/ft² and two years hence, a weight of 0.191 lb/ft² appears reasonable. This compares to the present weight of the deployable solar array of 0.449 lb/ft² (see Table 2).

Table 2. Solar cell module weights

Solar cell stack	Deployable solar array		JPL 30-W/lb array					
			State of the art		Optimum for probable advances			
		Lb/ft ²		Lb/ft ²	Within 1 yr	Lb/ft ²	Within 2 yr	Lb/ft ²
Coverglass	6 mil	0.074	6 mil	0.071	3 mil	0.036	3 mil	0.036
Coverglass bond	0.003 LTV-602	0.022	0.022 LTV-602	0.013	0.002 Sylgard 182	0.013	Integral	—
Solar cell	12.3 mil avg.	0.209	8 mil	0.092	8 mil	0.092	6 mil	0.069
Cell interconnections	Expanded silver	0.019	½ No. connect	0.010	Same	0.010	Same	0.010
Substrate bond	0.012 RTV-108	0.061	0.006 RTV-108	0.032	0.004 RTV-108	0.021	Same	0.021
Substrate	3-mil "H" Film	0.023	2-mil "H" film	0.015	1.5-mil "H" film	0.011	Same	0.011
Lateral stiffeners	6-.025 × 0.25 Mg	0.018	2-.025 × 0.25 Mg	0.007	Same	0.007		0.007
Foam	½ polyurethane	0.004	Same	0.004	Same	0.004		0.004
Foam bond	RTV-102	0.019		—		—		—
Ground wiring		—		0.010		0.010		0.010
Hinge pin		—		0.008		0.008		0.008
Electrical connectors		—		0.0001		0.0001		0.0001
Total weight/ft Wt (250 ft)		0.449		0.262 65.5		0.227 56.8		0.191 47.7

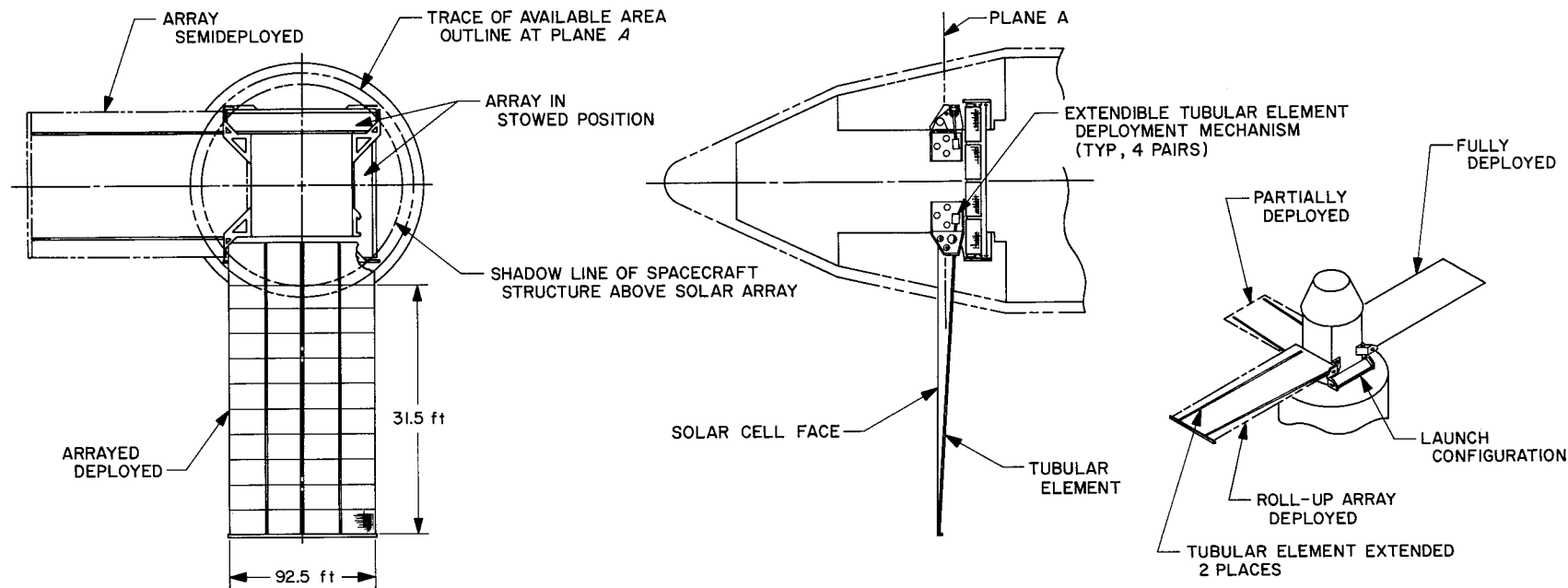


Fig. 3. Roll-up array tubular element design

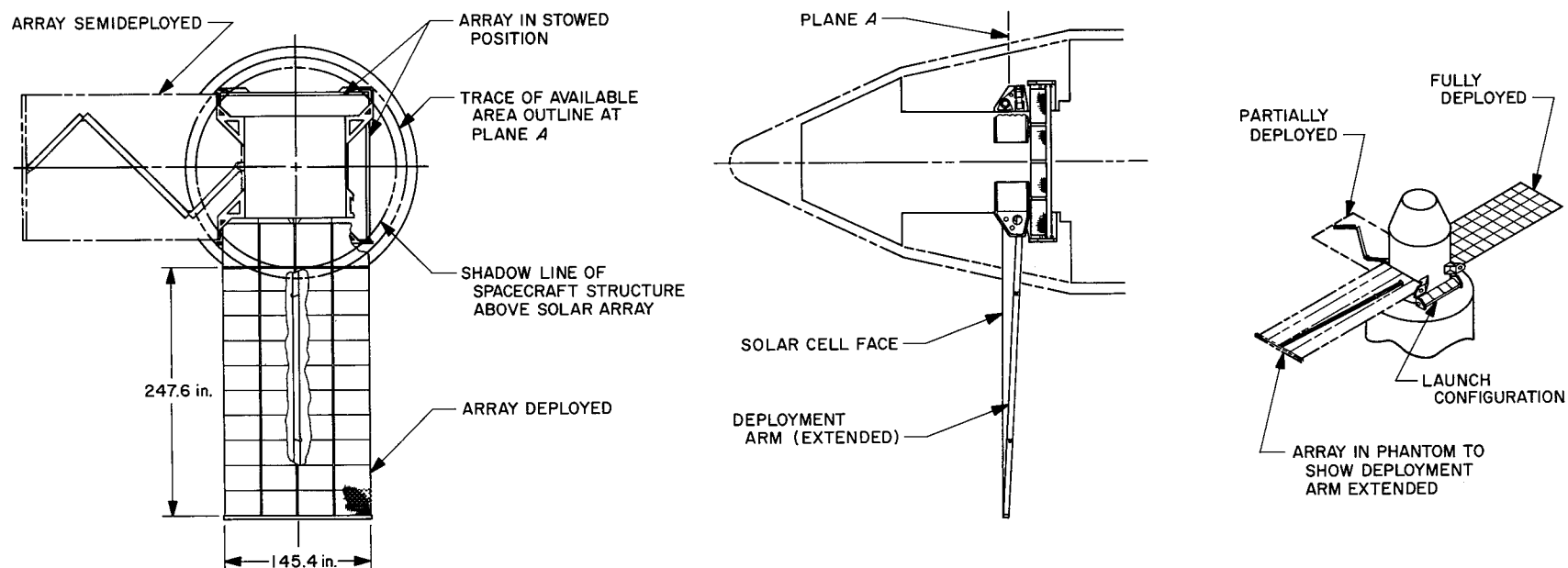


Fig. 4. Roll-up array deployment arm design

The mechanical linkage design of a panel shown in Fig. 4 is composed of three hinged sections mounted at the base of the spacecraft and deploying outward with a relative straight line motion at the tip. The arms are 4-in. aluminum tubing that extend a distance of 23.9 ft. The arms may be driven by an electric motor and/or may be mechanically driven by torsion springs.

The second Fairchild-Hiller design, requiring two tubular extendible element mechanisms to deploy the substrate from the roller outward to its final position, is shown in Fig. 3. A spreader bar is used between the tips of the booms for substrate attachment. A dual system is used to provide torsional capability by differential binding in the tubular elements. A mechanical or electrical synchronizing device will insure uniform deployment.

Study efforts are continuing in the areas of structural design, material selection, thermal and final selection of the deployment mechanism. Preliminary design studies are scheduled for completion late in December 1967.

General Electric Co. is proceeding along the lines of a single beam structure (Fig. 5) which will be either the Hunter Spring spiral wrap or the de Havilland storage tubular extendible member (STEM). At this time the de Havilland STEM is most likely to be chosen because of availability and a small weight advantage (Table 1). This single beam approach was chosen after evaluation of seven different arrangements because it best met the system requirements with the least complexity. In the single beam construction, torsional stability is controlled by the blanket tension; dynamic analysis is under way to define the effects and practicality of achieving the needed forces.

Substrate materials have been evaluated for specific gravity, tensile strength and elongation, shrinkage, tearing strength, folding endurance, coefficient of linear expansion, moisture absorption, dielectric strength, ultraviolet radiation, and electron, neutron, and gamma radiation. Among the materials reviewed were Teflon, Tedlar, nylon 6, Lexan, Mylar and Kapton. The material selected

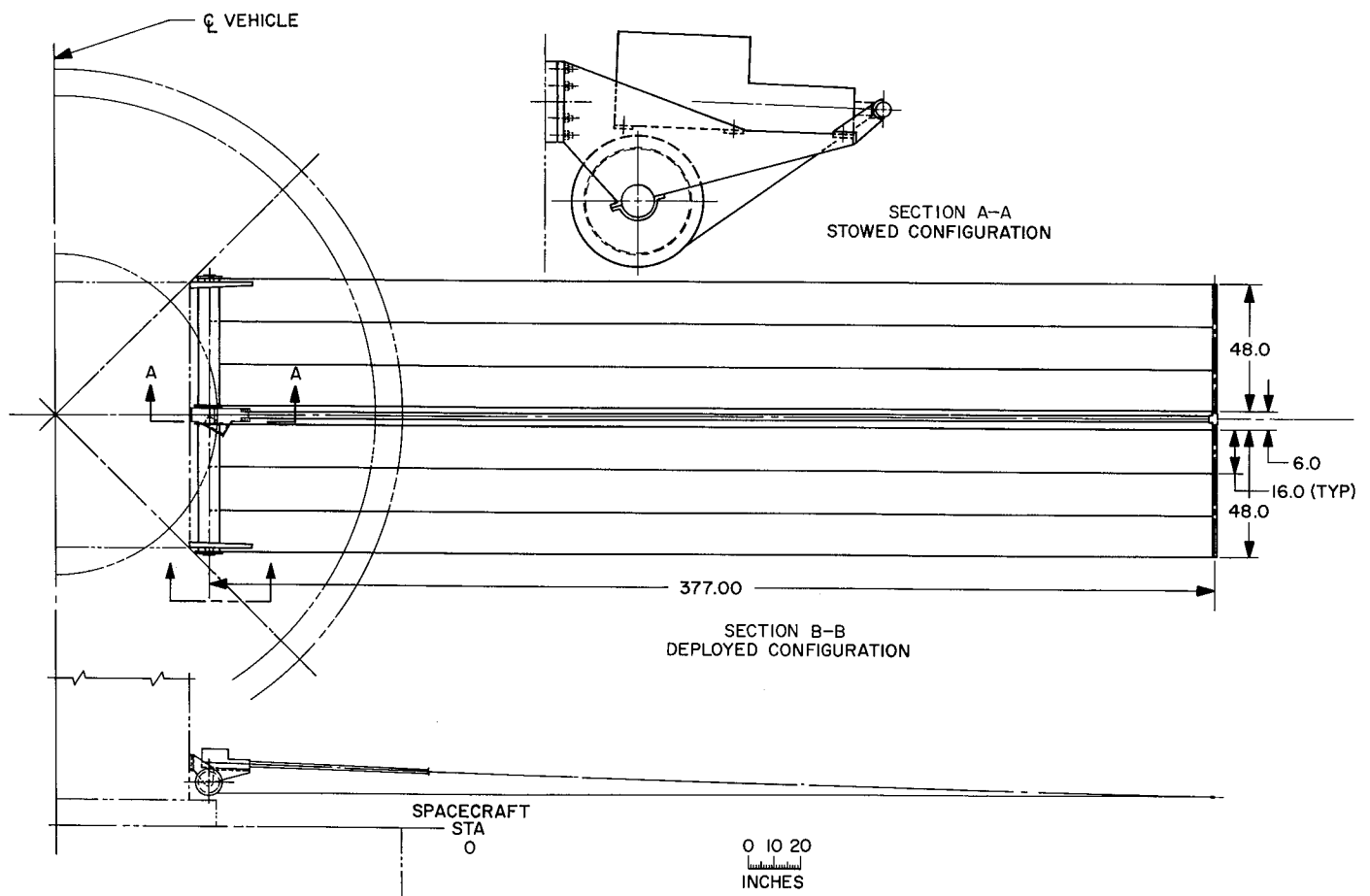


Fig. 5. Roll-up solar array single deployment rod showing modified drum supports (deployed configuration)

by General Electric as best suited for this design was Kapton.

D. Active Electronic Load, G. Stapfer

1. Introduction

The parametric performance testing of power generating devices such as thermionic or thermoelectric power converters requires some convenient means of simulating the power conditioning equipment normally used in a spacecraft power subsystem. In the past, this output load has been a passive device, e.g., resistors, water-cooled slidewires, etc. This article describes a highly stable and efficient electronic load which permits the recording of accurate and repeatable power sources performance data.

2. Purpose

In the performance testing of thermal power sources, it is imperative to record and maintain stabilized temperatures throughout the power source under test. Changes in the output voltage and current parameters will affect the amount of electron cooling experienced by the power source, thus modifying its temperature profile. This effect further amplifies the original voltage or current drifts, thus increasing the thermal instability of the power source. The recording of a single steady-state data point may thus become a very tedious and lengthy process.

It is the purpose of this newly developed electronic load to maintain a given output power parameter constant over a wide range of specified values. This will decrease the time required for the power source to reach thermal equilibrium, thus reducing the over-all testing time required. The recording of a complete volt-ampere curve (I - V curve) for a given temperature parameter is very time-consuming when plotted only with steady-state data points. To dynamically record these I - V curves, in the past, has usually resulted in a shift of this curve from its original steady-state value. The use of relatively fast sweep speeds in the electronic load has practically eliminated these offsets. This makes it possible to record the power sources' performance data more efficiently and accurately.

3. Description

A simplified block diagram of the electronic load is shown in Fig. 6. The load consists basically of a high-gain amplifier, a driver, a high current power transistor, and a feedback loop. For the sake of brevity, the sweep logic circuits are omitted from the block diagram.

The system exhibits an approximate open loop gain of 1×10^6 . The high gain will insure that the regulated output parameter will remain constant over a wide range of input perturbations. For example, in the constant voltage mode, the load output voltage will remain constant within ± 10 mV with currents ranging from a few milliamperes up to 200 A.

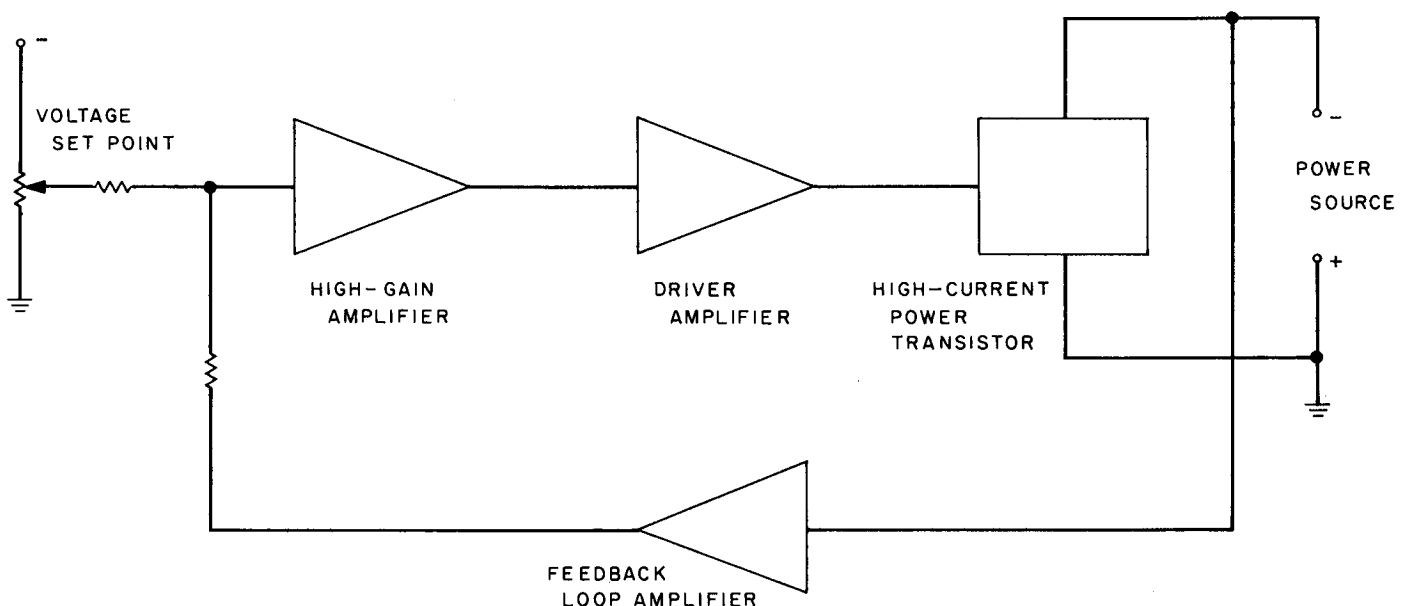


Fig. 6. Simplified block diagram of electronic load

To guard against any possible oscillations of the load, stemming from this high loop gain condition, the entire load is stabilized by the use of feedback capacitors.

In the constant current mode, a low-gain inverting amplifier is placed in the dc feedback loop, and the load current is sensed through a very low resistance current shunt ($0.1\text{ m}\Omega$). To improve the signal-to-noise ratio, the low-gain inverter is physically located near the load shunt. A high signal-to-noise ratio is important, since 1 mV of noise at the input of this inverter amplifier will appear as a 10-A sweep at the output of the load circuitry.

The main power transistors are placed as near as physically possible to the power source. This minimizes the voltage drops caused in the wiring and thus allows near short circuit conditions to be obtained. Testing a high current-low voltage power source such as a thermionic converter, this short circuit condition will typically be about 0.2 V at 200 A .

4. Operating Modes

The electronic load is capable of performing in any of four distinct modes of operation: constant voltage, constant voltage sweep, constant current, and constant current sweep.

a. Constant voltage mode. In this configuration the output voltage of the power source to be tested is adjusted for a desired output voltage. The load circuitry then will maintain this voltage output constant at this point, regardless of any other power sources' parameter changes. This will allow independent measurements to be performed on other variables of the power source, e.g., output current as a function of temperature, etc., at fixed output voltage levels.

b. Constant voltage sweep. A complete volt-ampere characteristic curve of the power source may be obtained while maintaining the load impedance at a constant voltage equivalent. This is accomplished by instantaneously short circuiting the power source, and then sweeping through the I - V characteristic to open circuit condition at a constant rate. A logic switching circuit will return the load terminal voltage to its previous quiescent setting. The sweep time from short circuit condition to open circuit voltage is adjustable from a few milliseconds to a full second. The type of power source, the recording equipment utilized, and the data accuracy desired will determine the optimum sweep speed utilized.

c. Constant current mode. In this mode the output current of the power source is adjusted for a given current output, and the load circuitry will maintain this current constant. As in the constant voltage mode, some other power source parameter may thus be determined without changing the output current.

d. Constant current sweep. The volt-ampere characteristic of the power source is obtained in this configuration while maintaining a quasi-constant current output. Its operation is the same as in the constant voltage sweep mode, except that the sweep will commence at the open circuit condition and sweep toward the short circuit point.

5. Conclusion

The electronic load described has been tested and operated on thermionic power converters. It was found to perform very satisfactorily in all modes of operation. As a direct result of utilizing this new load, the required time to completely performance-test a typical thermionic converter was reduced by approximately 50%.

E. Thermionic Development, P. Rouklove

The development of thermionic energy conversion systems is continuing at JPL, with emphasis on converter improvements and integration of converters with heat pipes. The heat pipe, a novel device, appears very promising as a means of thermal energy distribution.

1. Converter Tests

The tests of the converters of advanced technology are being accomplished at JPL. The series 9 converter most recently tested was built by Thermo-Electron Co., (TECo) Waltham, Mass. and was designated T-206; its general design is similar to that discussed in SPS 37-45, Vol. IV, pp. 31-39 and Fig. 7). Converter T-206 differed from the previous configurations in that the collector was capped with a thin rhenium sheet, pressure-bonded to the molybdenum substrate. The converter was tested at three values of emitter temperature, 1600 , 1700 , and 1800°C , recorded in a 8-to-1 ratio hohlraum. The tests were performed in a steady state condition for 5 values of voltage output for each thermal condition. Dynamic sweeps were taken at each steady state point of measurement. Emitter and collector apparent work functions were also measured for different emitter-reservoir and collector-reservoir temperature ratios (T_e/T_{cs} or T_c/T_{cs}) using the saturation and retarding plot techniques.

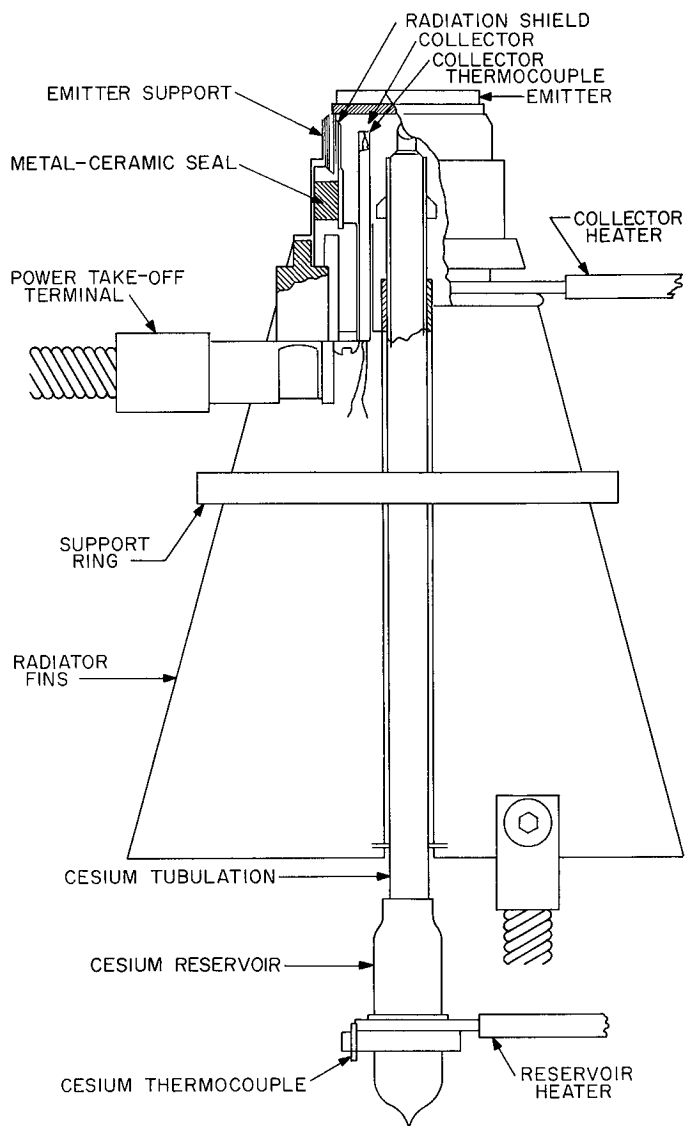


Fig. 7. Series 9 thermionic converter

The results of the parametric tests are graphically presented in Fig. 8. The power output of the converter did not achieve the expected values because of higher-than-desirable collector face temperatures. The geometry of the converter, with its long thin radiator fins, and large collector slug, results in a very high thermal drop (approximately 300°C) in the collector structure at high power outputs. Figure 9 presents comparative curves of converter T-206 and VIII-P-2, another prototype built in 1964 with 2 cm² of collector area. Both data are taken at emitter temperatures of 1700 and 1800°C while the collector temperatures shown were those observed at the various voltage output points. The power density in converter T-206 at 1800°C emitter temperature was 14.4 W/cm² as compared with 22.0 W/cm² demonstrated by

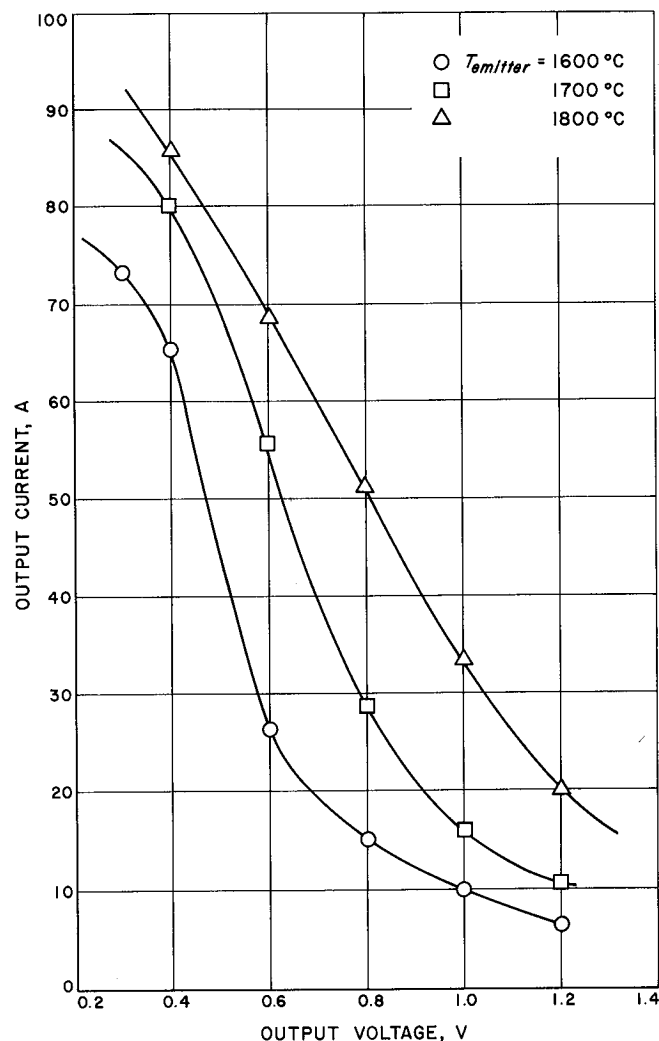


Fig. 8. Converter T-206 I-V characteristics

the older prototype. However, an improvement in the converter power output was observed when comparing the results of converter T-206 with those observed in converter T-205 (Fig. 10). This was attributed to a lower apparent collector work function observed in T-206.

The measurements of the emitter and collector work functions were performed assuming 2.750 cm² as the emitter surface area and 2.550 cm² as the collector surface area. The apparent emitter work function measurements were obtained from the saturation curves at T_e/T_c ratios of 4.15, 3.96, 3.76, 3.56, 3.36. The respective emitter work functions observed were 3.63, 3.45, 3.24, 3.04, and 2.85 eV, values slightly higher (by approximately 0.02 eV) than observed in converter T-205. Both emitters were solid rhenium electro-etched and thermally stabilized. The collector work functions were obtained by retarding plot techniques at T_c/T_e values of 1.47, 1.45, 1.41, and 1.37.

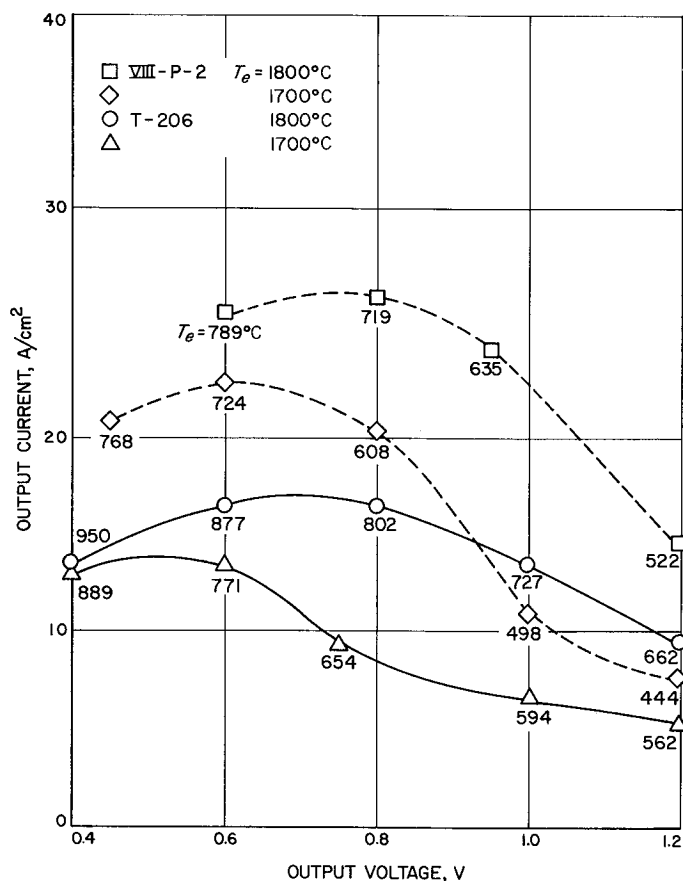


Fig. 9. Converter current output density versus emitter and collector temperature

The respective apparent collector work functions observed were 1.47, 1.44, 1.49, and 1.50 eV. These values are approximately 0.50 eV lower than those observed in T-205. No measurements of the interelectrode spacing were performed on T-206 due to the absence of a collector heater and the resulting drift in the reference point with changes in power input values. From the foregoing it has become apparent that no significant improvement in power output can be expected with this configuration, the radiator geometry being the limiting factor.

2. Heat Pipe

The advantages and operating principle of a heat pipe were presented in SPS 37-45. This effort is being pursued at JPL with two parallel applications. The effort at TECo covers the use of the heat pipe to replace the conventional radiators as a means for heat rejection from the collector. Under contract with Radio Corporation of America, Lancaster, Pa., the effort is devoted to supplying heat to the emitters of the thermionic converters. The goal of

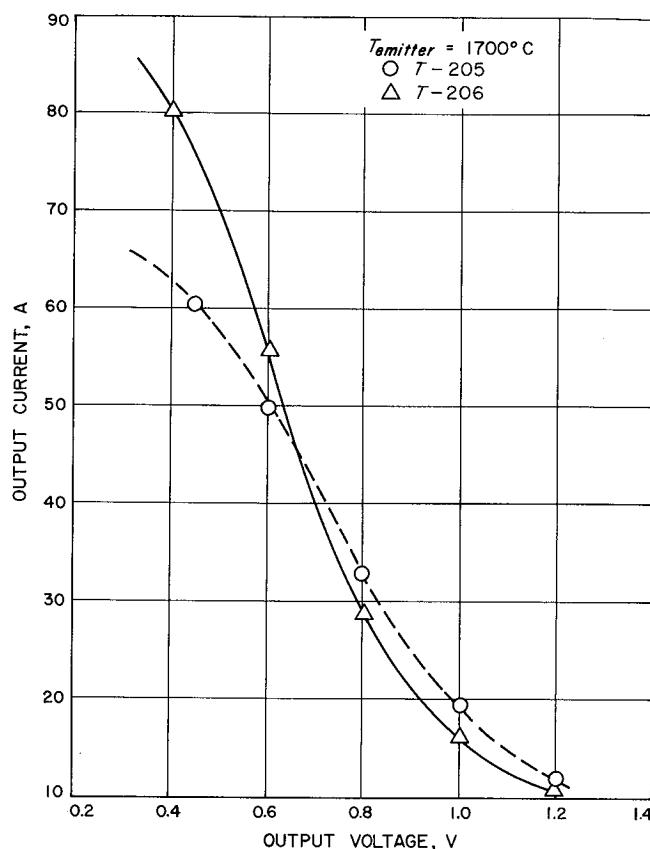


Fig. 10. Converters T-205 and T-206 comparative I-V characteristics

these efforts is the integration of thermionic energy conversion with a thermal source, especially isotopes.

After several iterations, during which multiple compatibility problems were solved, a heat pipe was successfully integrated with a collector structure of the series IX converters (Fig. 11, converter T-3). The walls of the heat pipe, covered by a chromium oxide high-emissivity coating, act as a radiator with 37.5 cm² area. The converter has an integral niobium collector and heat pipe structure. The converter performed successfully, but not as well as was anticipated because of the higher apparent work function of the collector material (Ref. 1). The converter also successfully passed 12 consecutive cycling tests from room temperature to full power and performed for 400 hr. Unfortunately a material deficiency in the cesium supply tubulation created a cesium leak which could not be corrected.

Figure 12 presents comparisons of the collector temperatures of the converter with a heat pipe to similar converters with conventional fin radiators. An increase in heat pipe diameter (T-4) was found desirable. This involves

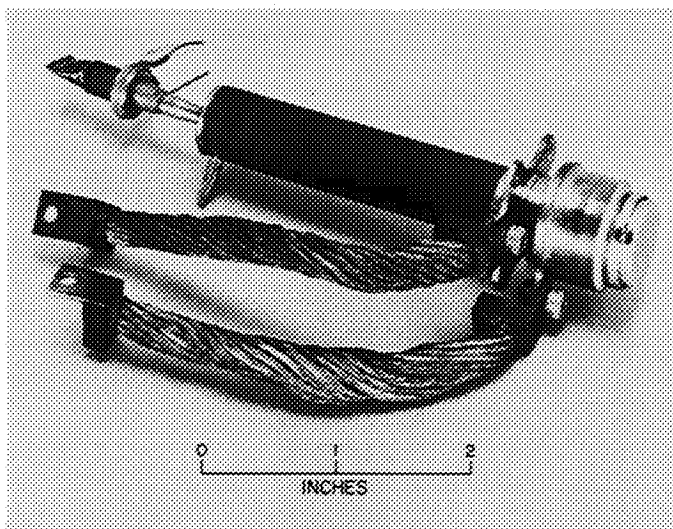


Fig. 11. Converter with heat pipe collector (T-3)

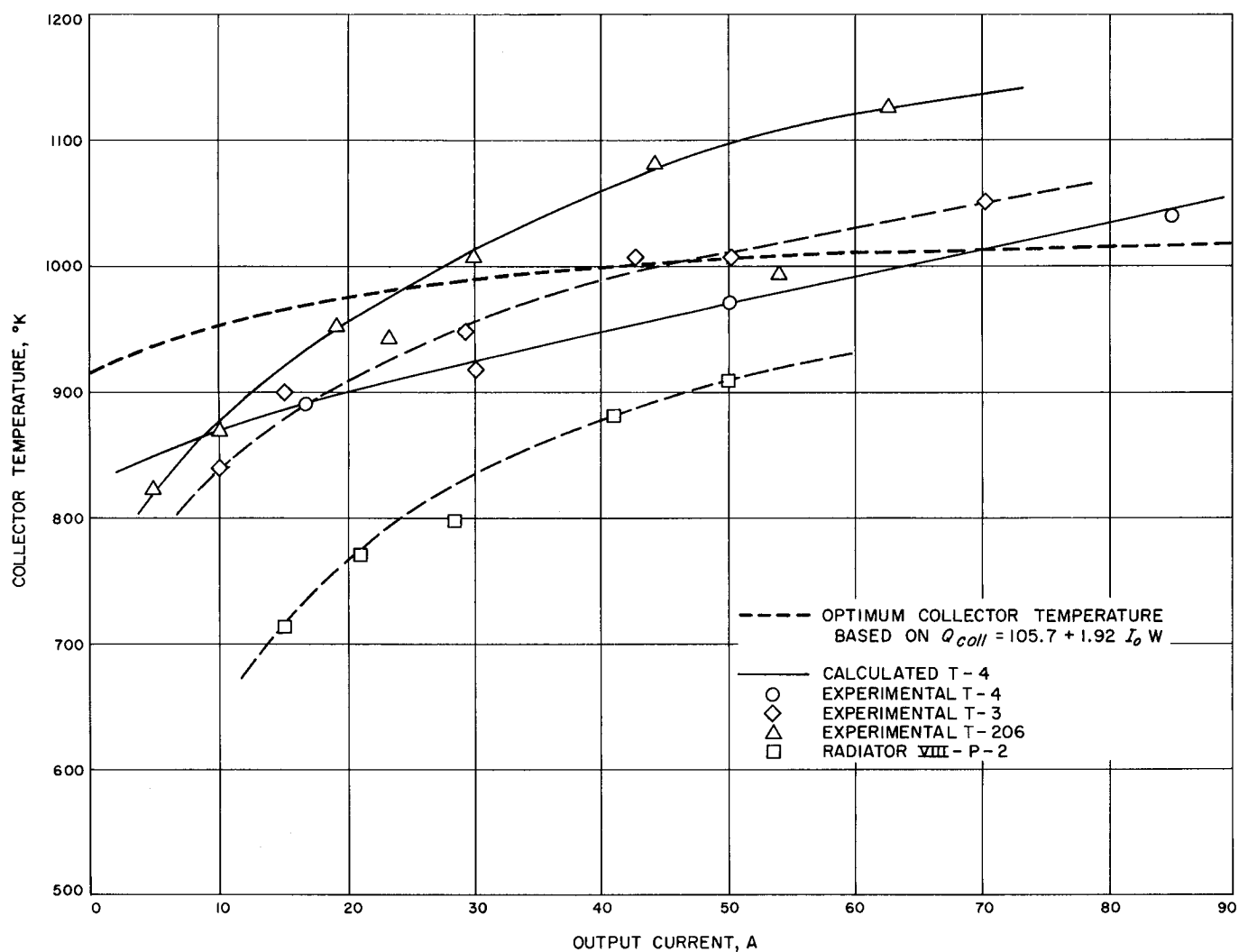


Fig. 12. Collector temperature versus current output

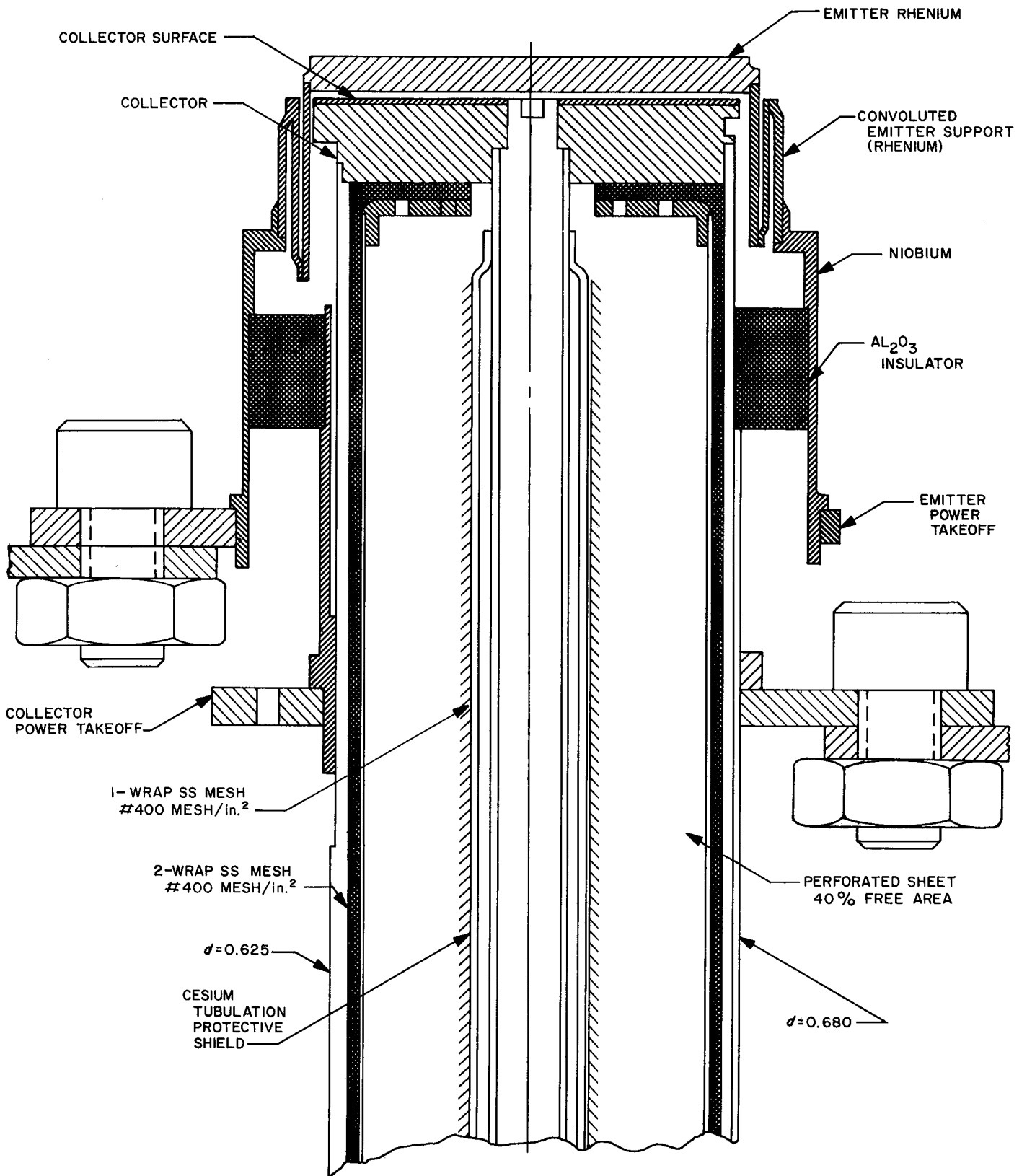


Fig. 13. Collector heat pipe designs

a modification of the seal structure (Fig. 13, right half) in which the alumina insulator is brazed directly to the heat pipe wall and not to an additional member as previously (Fig. 13, left half). Experiments performed on one such structure indicated that this modification did not present a problem. However, during the thermal cycling tests, excessive heat from the electron bombardment suddenly concentrated on the relatively small collector area provoked nucleated boiling, which resulted in the "drying" of the capillary and the destruction of the heat pipe assembly by the melting of the stainless steel wick and its subsequent alloying with the niobium container. It is not expected that a similar defect will occur during normal converter operation because of more gradual and uniform heating.

The emitter heat pipe previously under long-term life test failed after 10,526 hr of consecutive operation at temperatures ranging between 1500 and 1600°C. This heat pipe was built of TZM, used molten lithium as a working fluid, and during operation was transferring 2 kW thermal. The failure was traced to a closure weld which previously had shown to be a weak point. All welds are now X-rayed before final closure of the pipes. The successful demonstration of long-term operation of a heat

pipe at high temperatures gives hope to the possibility of building a highly reliable multiconverter thermionic generator using the heat pipe as means of distributing heat to the emitters. The isothermal properties of the heat pipe and its capability of thermal flux concentration (SPS 37-45) render it especially useful for radioisotope-heated power systems.

In an effort to increase the thermal conduction capability of the pipes, several capillary configurations were tested (channel, mesh, corrugated sheet, and combinations of several of these). None of these arrangements could transfer more than 4 kW(t) without "hot spots" and uneven circumferential heat distribution. This limitation, now under investigation, appears at least partially to be the result of heavy magnetic fields created by the pipe-heating method (resistance heating). The magnetic fields could constrain the liquid metal flow to the inner part of the capillary structure, separating it from the walls and thus creating "hot spots."

Reference

1. Brosens, P. J., "Advanced Converter Development," IEEE Thermionic Specialist Conference, Palo Alto, Calif., Oct. 30-Nov. 1, 1967.

VI. Guidance and Control Analysis and Integration

GUIDANCE AND CONTROL DIVISION

A. Capsule System Advanced Development Operational Support Equipment, K. Mussen

Three sets of operational support equipment are being built to support the capsule system advanced development program. To reduce costs and to expedite delivery, the OSE is being fabricated by modifying surplus *Ranger* and *Mariner* OSE.

The three flight subsystems for which OSE is being built are the lander power subsystem, entry power subsystem, and the lander sequencer and timer. An additional piece of hardware is an entry sequencer and timer subsystem switching panel which simulates the command functions normally provided by an ES&T. This ES&T panel is included with the EPS OSE.

Both sets of power subsystem OSE use surplus *Ranger* equipment, with new control panels to provide the new logic. The LS&T OSE is modified *Mariner* Venus 67 OSE.

The LPS OSE description typifies the OSE requirements and the methods used.

1. Lander Power Subsystem OSE

The LPS OSE is contained in two 5-ft-high racks; one rack contains the controls, command logic, dummy loads,

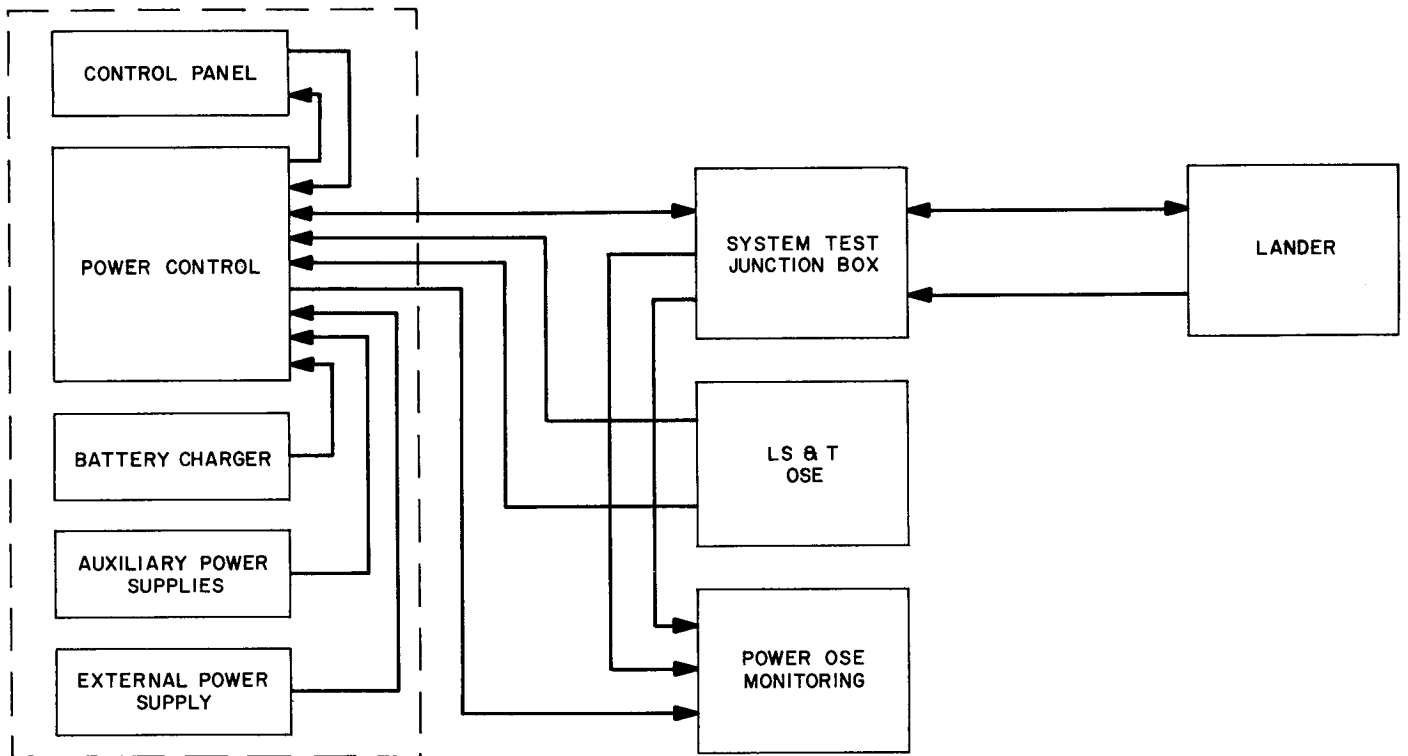
battery charger, and power sources, while the other rack contains the monitoring equipment. The OSE design permits the same two racks to be used both for laboratory tests of the subsystem alone and for support of system tests where the power subsystem is installed in the capsule.

The differences in configuration between laboratory test and system tests operations are shown in Fig. 1. Different junction boxes are used, and during a laboratory test, the control panel simulates capsule commands that are normally generated by the LS&T.

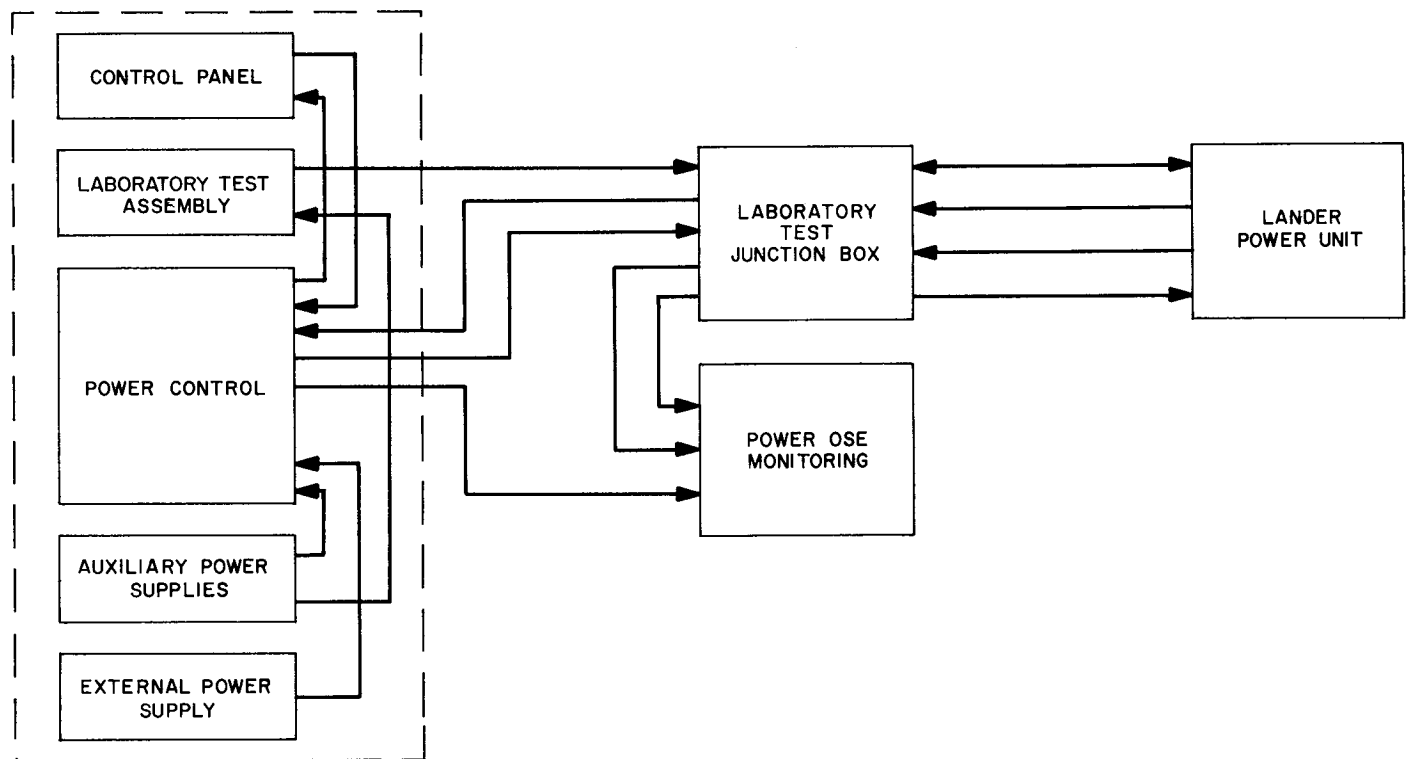
The functions required of the lander power OSE include: supply capsule external power; control all power functions; supply simulated commands (laboratory test only); provide loads for the power system (laboratory test only); provide monitoring of all required functions; and charge the lander battery when required.

While performing the above functions, the power OSE also must protect the capsule from operator and equipment failures.

Protection of the capsule from accidental shorts or overloads is accomplished by: overvoltage interlocks; overvoltage zener clamp; current limiting external power source; current limiting resistors; and power control interlocks.



SYSTEM TEST CONFIGURATION



LABORATORY TEST CONFIGURATION

Fig. 1. Capsule system advanced development lander power system

To provide additional safety, the LS&T OSE and power OSE are interlocked to prevent application of power to the lander capsule if the LS&T is not in a proper turn-on condition. Another interlock, "pyro unsafe," is provided; in the event of an unsafe pyro condition, it will transfer the lander capsule to external power, and then turn the capsule power off automatically.

The monitoring rack consists of a modified *Ranger* Block III digital data acquisition system plus a battery cell voltage monitor meter (cell voltages are also recorded in the data acquisition system). All data channels have permanently-installed RC filter networks to prevent voltmeter filter charging transients occurring as different channels are selected.

Observation of pyro events is accomplished by monitoring the voltage appearing across simulated "burn wires" which are located in the junction box. The high speed recorder, used for monitoring these events, is located in the pyro OSE.

Battery charging is accomplished with a commercial 0-1 A constant current power source. Stop-charge is performed manually.

Fabrication of the lander power subsystem OSE is well under way; four assemblies have been fabricated (or modified) and tested, and the intra-rack cabling is completed.

2. Entry Power Subsystem OSE

The EPS OSE is contained in two 5-ft-high racks; one rack contains the reference power supplies and the assemblies for monitor and command functions. The OSE design permits the same two racks to be used both for laboratory tests and for system tests.

Design and fabrication are well under way with seven assemblies complete, two assemblies requiring minor modifications, and two assemblies with design and fabrication in process. The intra-rack cabling is nearly completed. Fabrication of the ES&T simulator panel has been started.

3. Lander Sequencer and Timer OSE

The LS&T OSE is contained in one 6-ft-high rack containing the assemblies to be used both for laboratory tests and system tests. A signal conditioning box is also required to condition the signals passing to and from the OSE and LS&T.

Design is nearly complete and fabrication is well under way. Four assemblies are complete and two are being modified. The intra-rack cabling is nearly complete, but the design and fabrication of the signal conditioning box have not been started.

B. Gas Valve Flow Detector, S. D. Moss

During performance testing of a spacecraft, it is necessary to ascertain the actual operational periods of the attitude control nitrogen gas jets. Each spacecraft employs a total of 12 of these jets, located on the solar panel extremities, for spacecraft stabilization and pointing. The present method utilized to sense the operation of these gas jets consists of four junction boxes located remotely from the spacecraft at distances which are dictated by the particular test requirements. These junction boxes contain either two or four diaphragm-type pressure switches mounted on a gas-tight manifold, with associated check valves. Fittings are provided for tygon tubing connections to each spacecraft gas jet, check valve, external vacuum roughing pump, and self-test gas source. Provision is also made for electrical cabling connections between the junction boxes and the attitude control operational support equipment (OSE). Some of the undesirable characteristics of the present mechanization are:

- (1) Unsatisfactory pressure switch operation (slow response time, low reliability and questionable environmental characteristics)
- (2) Excessive system complexity
- (3) Slow system response (due to long lengths of tygon tubing)

In view of the disadvantages of the present mechanization, the development of a gas valve flow detector to overcome the described limitations was initiated. A thorough background study of the present mechanization was completed and the inadequacies of the existing sensor mechanization ascertained. The results of this study, in conjunction with information obtained from existing flight and OSE environmental specifications, were used to establish a meaningful design specification for the new detector.

The over-all performance of the gas valve flow detector is determined primarily by the transducer characteristics since considerable flexibility exists for the design of the associated signal conditioning electronics. Therefore, the

selection of the appropriate transducer is of prime importance.

The transducer types investigated were: (1) low time constant thermistors, (2) piezoelectric ceramics, (3) solid state strain gauges, and (4) hot wire anemometers.

Low time-constant thermistors have been eliminated because none of the commercially available devices has a sufficiently small size to yield the short time constant required.

1. Piezoelectric Detector

One of the more promising transducer materials for use in the gas valve flow detector is a piezoelectric ceramic which responds to gas turbulence-induced noise. Several Clevite Corp. PZT-5 piezoelectric ceramic devices were produced for evaluation and testing. These devices included (1) a 0.250-in. diam, 0.100-in.-thick disk, with a 0.050-in. orifice and (2) a 0.250-in.-long tube having an outside diameter of 0.250 in. and a thickness of 0.020 in., as shown in Figs. 2 and 3, respectively. As a design aid, the frequency characteristics of the gas flowing from the gas valve were desired. These were obtained by connect-

ing a spacecraft gas valve to a dry nitrogen gas supply and exciting the valve solenoid with a special pulse circuit developed for sensor evaluation. The audio-frequency characteristics of the gas flow from the valve nozzle over the range of 100 Hz to 200 kHz were recorded for several nozzle orifice sizes and for selected sensor locations. These data were obtained utilizing a precision condenser microphone and a panoramic sonic analyzer.

The disk and tube ceramics were tested in various locations with respect to the valve nozzle. Due to the small signal level associated with these devices, a differential amplifier circuit was designed around a Fairchild-type μA 709 monolithic operational amplifier.

The disk was tested by utilizing a 1½-in. length of heat shrinkable tubing between the valve and the disk. The disk was oriented perpendicular to the gas flow and a seal effected around the disk such that the total flow of gas was discharged through the orifice. The ceramic tube was tested in a similar way by orienting the tube such that the gas was directed longitudinally through its center. A frequency spectrum analysis of the transducer output signal revealed random frequency content above 500 Hz.

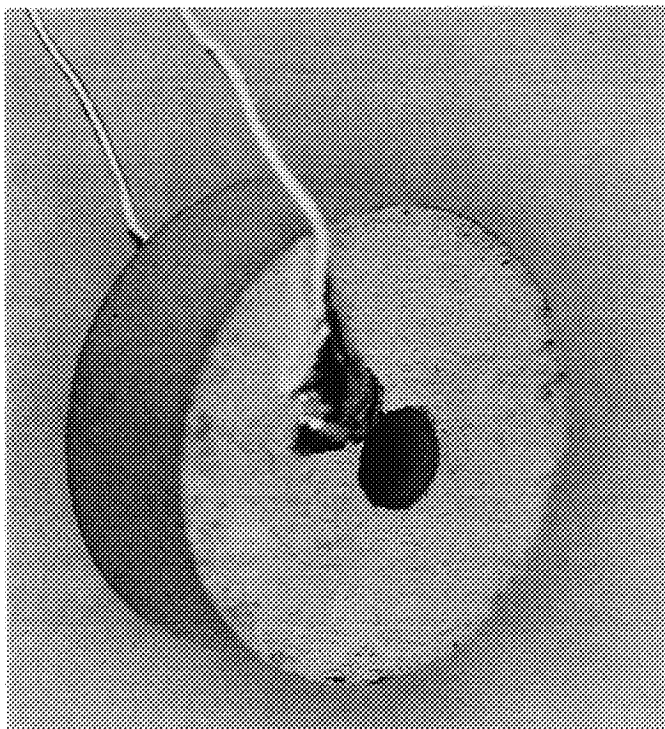


Fig. 2. Piezoelectric ceramic sensor disk

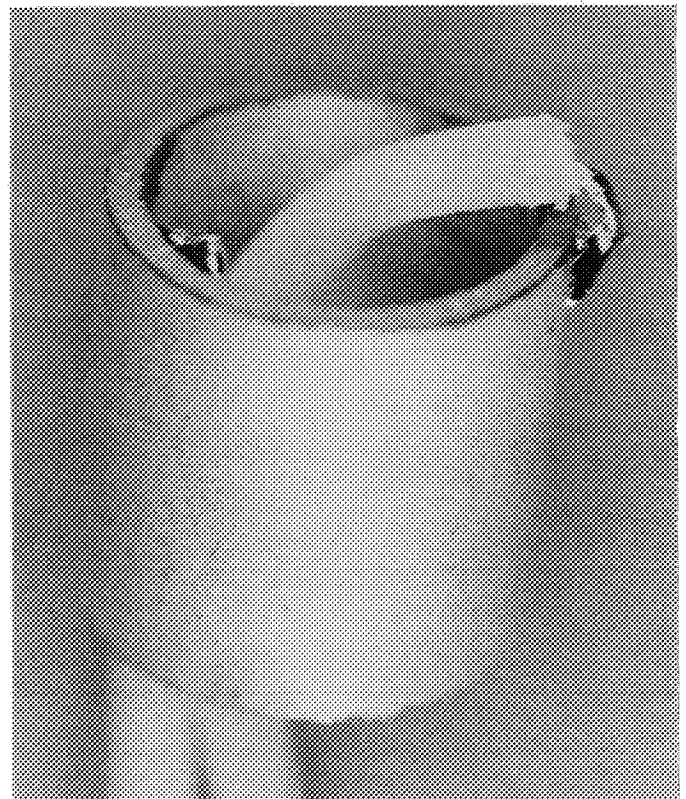


Fig. 3. Piezoelectric ceramic sensor tube

For the ceramic transducer to be useful, the nearly white noise output of the ceramic element must be processed through a signal conditioner to derive a dc pulse corresponding to the open duration of the valve. The standard integrating-type charge amplifier commonly used with ceramic transducers is not necessarily a wise choice for the signal conditioner, because the low frequency pass band is too broad to adequately limit external noise. A better method is to convert the differentially connected monolithic operational amplifier into a very

narrow active band-pass filter. A computer program was developed for the synthesis of the input and feedback networks associated with active low-pass, high-pass, and band-pass filters to aid in the design of the signal conditioning electronics. A prototype detector will be fabricated and tested in the near future, utilizing the disk piezoelectric ceramic sensor. The signal conditioning electronics are presently being designed to produce a nominal 10-V dc output pulse which will correspond to the duration of the 20-msec random noise signal produced by the ceramic transducer.

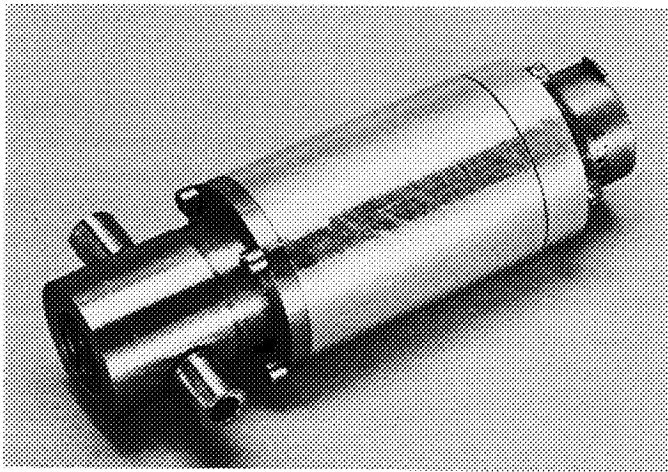


Fig. 4. Prototype gas valve flow detector

2. Hot Wire Detector

A second transducer having favorable characteristics is a hot wire anemometer consisting of two 0.00015-in.-diam platinum (containing 10% rhodium) filaments mounted on a TO-5 integrated circuit header. One filament, the active element, is positioned directly in the path of the nitrogen gas flow from the gas valve. The second filament, the passive element, is located on the header in such a manner that it is shielded from the gas flow. Since both the active and passive elements are physically located in close proximity, they are subjected to approximately the same ambient temperature variations. This temperature tracking property is useful in that the passive element provides the necessary temperature compensation for the flow sensor when connected in a bridge configuration.

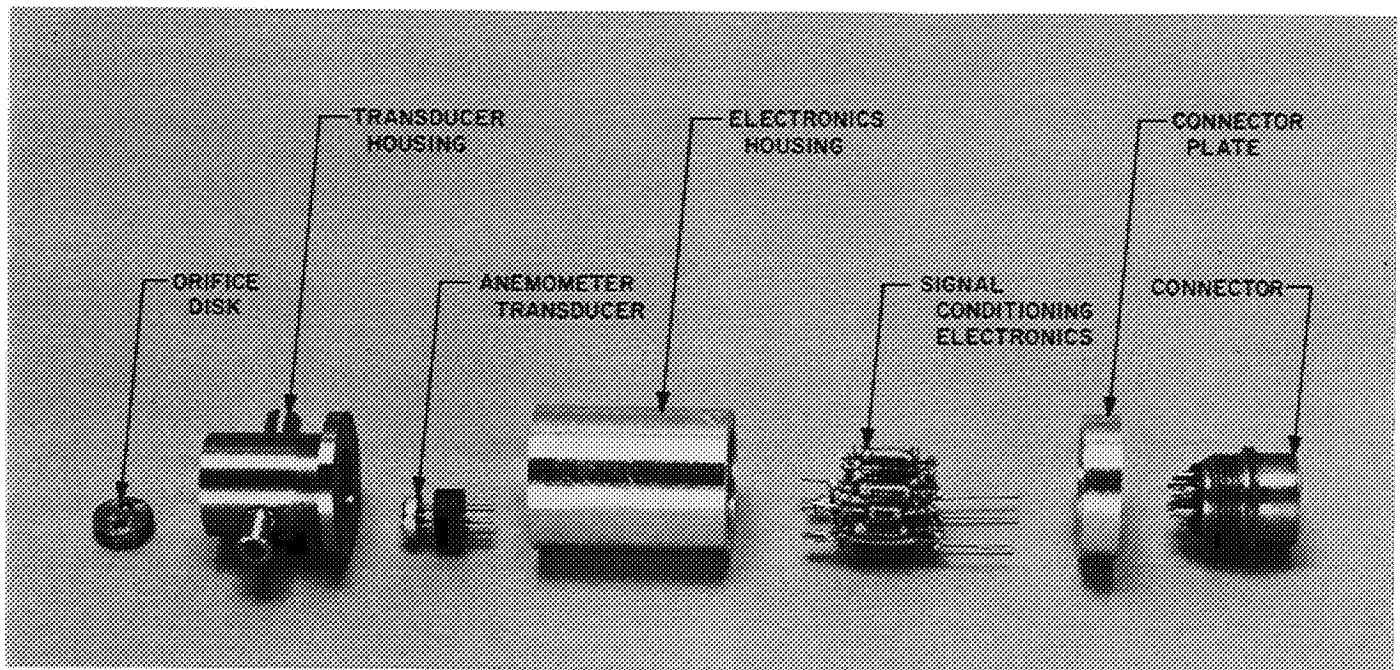


Fig. 5. Prototype gas valve flow detector, exploded view

Unlike the signal conditioning requirements associated with the piezoelectric ceramics, a differentially connected monolithic operational amplifier is sufficient for this application.

A prototype gas valve flow-detector utilizing a hot wire anemometer has been designed and fabricated for evaluation testing. This prototype, as shown in Fig. 4, is approximately 3-in. in length and 1-in. in diameter. An exploded view of the prototype detector is shown in Fig. 5. Figure 6 shows the detector coupled to a typical *Mariner* spacecraft attitude-control gas valve.

The prototype detector has undergone various phases of acceptance testing. The smallest and largest filaments used in these tests have been 0.0001- and 0.0005-in. diam, respectively. The 0.00015-in.-diam filament selection resulted from a compromise involving signal response times and mechanical rigidity. The prototype detector was satisfactorily tested over an environmental temperature range from 0 to +185°F. At vacuum pressures of 10^{-7} torr, temperatures from 75 to 215°F have little effect on the operation of the transducer.

Some difficulty was experienced in testing the unit at vacuum pressures in the range of 10^{-5} to 10^{-7} torr. Due to the absence of atmospheric cooling, it was found that the platinum filament reached incandescence at a lower current as compared with ambient conditions. The vacuum bias current required for incandescence at 10^{-5} torr was determined to be approximately one-tenth that required for incandescence under ambient pressure conditions. One of two choices must be made for the anemometer to perform satisfactorily over a wide environmental pressure range. The first choice is to operate the anemometer at

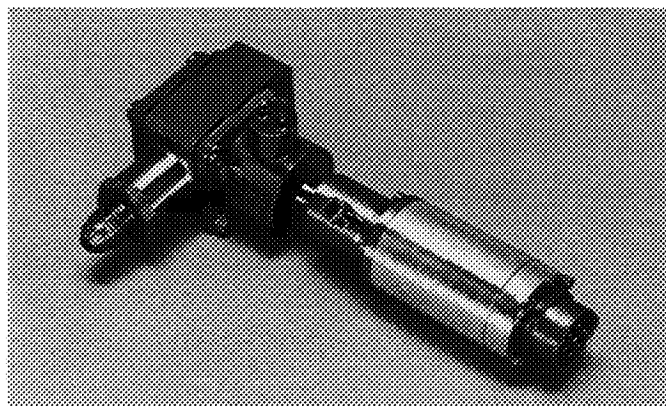


Fig. 6. Prototype gas valve flow detector connected to *Mariner*-type gas valve

a bias current dictated by the highest specified vacuum. The second choice is to derive a bias current which is a function of pressure. The first choice appeared to be the simpler of the two. However, the loss of sensitivity, resulting from the anemometer operating at such a low bias current when subjected to atmospheric pressure conditions, makes this choice unattractive. The second choice, a pressure-dependent bias current, was selected to solve the problem. The same characteristic which produced the problem, that of variation in thermal conductivity as a function of pressure, was utilized in an attempt to solve it. Different combinations of series sensistors and shunt thermistors were used in an effort to obtain a 10:1 compensation in bias current over the full pressure range. The maximum compensation obtainable utilizing this method was approximately 5:1. A solution to this problem has not been obtained to date, but may be a compromise between the two alternatives.

VII. Guidance and Control Research

GUIDANCE AND CONTROL DIVISION

A. Preparation and Physical Properties of α -Se,

S. Iizima and M-A. Nicolet¹

1. Introduction

Four distinct allotropic modifications of solid Se are known to date: amorphous, trigonal (also called "hexagonal" or "metallic"), α -monoclinic and β -monoclinic.

Se of high purity, traditionally obtained by vacuum distillation, is amorphous and commercially available to 99.9999% purity. The electrical properties of this material are strongly affected by traces of oxygen and halogens. Details on this subject are found in Refs. 1 and 2.

The trigonal modification has been studied actively in recent years. It is the stable form of Se at room temperature (Ref. 1) and is obtained readily in thin films by vacuum evaporation. The structure consists of spiral chains of Se atoms arranged in trigonal symmetry. The direction of rotation in the spiral can be either left- or right-handed. Stuke (Ref. 3) gives an up-to-date review of the present understanding of the charge transport mechanisms in this solid.

In contrast to this, little is known about the properties of the monoclinic modifications. The crystal structure has been determined to consist of rings with eight Se atoms as the basic building block arranged in slightly different patterns in the two modifications (Refs. 4-7). To the knowledge of the authors, only one paper gives data on the electrical transfer properties of monoclinic Se (Ref. 8).

We have decided to obtain more experimental information on monoclinic Se. By its structure this solid is representative of molecular crystals about which less is known than for covalent semiconductors. By its chemical composition the crystal is elemental and should therefore be easier to obtain in high purity and perfection than, for example, organic molecular crystals. Selenium has always played a dominant role in applications (rectifiers, photoelectric devices, xerography). These arguments all suggest that selenium should be a particularly rewarding material to study. Its merits as a potential carrier of space-charge-limited current will receive special attention.

2. Crystal Preparation

Single crystals of monoclinic Se are usually obtained by evaporating a CS₂ solution saturated with Se. Crystals

¹At the California Institute of Technology, performing work supported by JPL.

obtained in this manner are of either alpha or beta type and are always of submillimeter size. Kyropoulos (Ref. 9) prepared somewhat larger crystals in a vessel with a temperature gradient. Se is believed to be transported through the liquid phase by diffusion and convection in such a system. By this method, Kyropoulos obtained crystals 2 to 3 mm in size over a period of a few months. Larger crystals have been grown faster by a modification of this method. Crystals 2 mm in size are obtained in 2 to 4 wk. The apparatus is shown schematically in Fig. 1. Equivalent results are achieved in both fully sealed glass tubes and in glass tubes closed with a Teflon lid. The latter arrangement permits seeding with small crystals in the low temperature region. This significantly increases the yield of large crystals. Amorphous Se in pellet form of two different origins has been used as initial material with comparable success.²

3. Physical Properties

The habit of our crystals is equal to that described by Mitscherlich (Ref. 10). Figure 2a shows a typical form and the nomenclature introduced by this author for the various surfaces. According to Burbank (Ref. 5), the crystallographic assignment of the p-face is (101) in the reference system used by him (Fig. 2b).

The density of these crystals has been determined; a random selection was ground to powder of a total weight of approximately 200 mg. The weight of powder in a small glass bulb and the weight of the glass bulb alone (approximately 300 mg) were measured in air and in a liquid of known density to an accuracy of ± 0.2 mg. The result of such measurements is shown below:

Temperature, °C	Density, g/cm ³	Liquid used
20.8	4.26 ± 0.075	Benzene
18.5	4.26 ± 0.05	Benzene
16.3	4.36 ± 0.05	Methyl alcohol
16.0	4.31 ± 0.05	Methyl alcohol

From the lattice constants reported by Burbank and the atomic weight of 78.96 for Se, a value of 4.38 g cm^{-3} is computed. The slight discrepancy may be due to residual voids in the liquid or the powder, although efforts were made to eliminate both by ultrasonic shaking of the liquid-powder mixture and fine grinding of the crystals. Similar measurements on the amorphous solid used as a starting material yielded a density of $4.03 \pm 0.02 \text{ g cm}^{-3}$.

²Canadian Copper Refining Ltd., Hyperpure Se; Gallard-Schlesinger Co., 99.9999% Se.

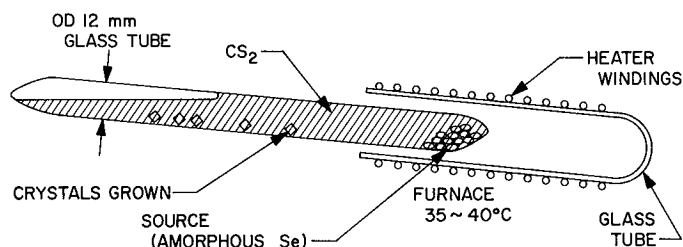


Fig. 1. Apparatus for growing α -monoclinic Se crystals

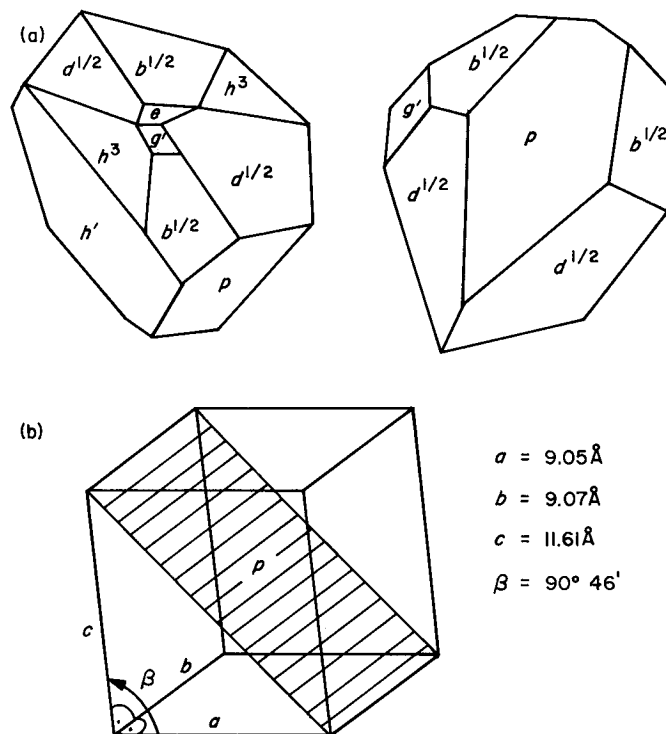


Fig. 2. (a) Typical habits of crystals, (b) Lattice structure of α -monoclinic crystal

A comparison of these data with the latest quotations in the literature is given in Table 1.

To determine the dielectric constant, some crystals were polished down along p-faces with $0.3\text{-}\mu\text{m}$ alumina on microcloth. Au films were evaporated onto both sides and the rims were removed by cleaving. The capacitance of the structures was measured with a Boonton 75C bridge. The results were consistent for frequencies from 5 to 50 kHz and for dc biases between ± 100 V. The area A (typically 10^{-3} cm^2) and the thickness d (typically $50 \text{ }\mu\text{m}$) of the samples were determined from microscopic enlargements and micrometer measurements. The average value obtained from several samples and the relationship $\epsilon = dC/\epsilon_0 A$ is $\epsilon = 6.5 \pm 0.6$. The large error reflects the

Table 1. Density of dielectric constant of Se in its different crystallographic modifications

Modification	Density, g/cm ³		Dielectric constant	
	Literature	Measured	Literature	Measured
Amorphous	—	4.03 ± 0.02	6.3	—
Trigonal	4.80 ^a	—	6.6 ^a	—
	4.86 ^b	—	6.37 ^c	—
α -monoclinic	4.46 ^b	4.30 ± 0.05	7.39 ^d	6.5 ± 0.6
	4.38 ^e			
β -monoclinic	4.42 ^b	—	—	—

^aRef. 1.
^bSnead, *Comprehensive Inorganic Chemistry*, Vol. VIII.
^cMoss, T. S., *Optical Properties of Semiconductors*, Butterworth, 1961.
^dRef. 9.
^eCalculated with Burbank's lattice parameters.

uncertainty in sample thickness due to nonparallel surfaces. Efforts are currently under way to improve processing techniques of these crystals and to achieve better control of sample geometry.

4. Concluding Remarks

Relatively large single crystals of α -Se have been successfully grown by a technique of mass transport in CS₂. Values for the density and the dielectric constant have been obtained which are not in full agreement with the literature. More accurate measurements will be possible when larger amounts of single crystals are grown and better processing techniques have been developed. Very thin and small platelets, apparently of the same crystal structure, have been obtained occasionally. Attempts are made to determine the relevant parameters controlling the growth of this crystal habit.

References

1. Hogarth, C. A., *Materials Used in Semiconductor Devices*, Interscience Publishers p. 49, 1965.
2. Eckart, F., *Ann. Phys.* Vol. 17, Series 6, p. 84 1955/56.
3. Stuke, J., *Festkörperprobleme* Vol. 5, p. 111, Vieweg and Sohn, Braunschweig, 1966.
4. Klug, H. P., *Z. Kristallogr.* Vol. 88, p. 128, 1934.
5. Burbank, R. D., *Acta Cryst.* Vol. 4, p. 140, 1951.
6. Burbank, R. D., *Acta Cryst.* Vol. 5, p. 236, 1952.
7. Marsh, R. M., Pauling, L., and McCullough, J. D., *Acta Cryst.* Vol. 6, p. 71, 1953.
8. Gudden, B., and Pohl, R., *Z. f. Physik.* Vol. 35, p. 243, 1926.
9. Kyropoulos, S., *Z. f. Physik.* Vol. 40, p. 618, 1927.
10. Mitscherlich, M., *Ann. de Chimie et de Physique*, Vol. 46, p. 301, 1856.

B. Magneto-Optic Information Storage, G. Lewicki

The objective of this work is to demonstrate the feasibility of using an optical beam to store and retrieve information on, and from, thin ferromagnetic films having an easy axis of magnetization perpendicular to the plane of the film and a large Faraday effect.

In the method of magneto-optic information storage being considered, small areas of the film correspond to bits of information. The state of a bit corresponds to the area being magnetized in one of two directions perpendicular to the plane of the film. Setting the magnetization within a bit in a desired direction, or storing information, is accomplished with a method known as Curie-point writing. A bit is heated past its Curie temperature where it becomes nonmagnetic and then it is allowed to cool in an applied magnetic field having sufficient intensity to set the magnetization within the bit in a desired direction. A focused laser beam can be used to heat bits.

Some theoretical considerations concerning the Curie-point-writing process have been presented in SPS 37-42, Vol. IV, pp. 59-61, and SPS 37-46, Vol. IV, pp. 84-87. The Faraday effect is used to determine the state of a bit, or the direction of the magnetization within an area. On passing through a bit, plane-polarized light has its plane of polarization rotated clockwise or counterclockwise, depending on the direction of the magnetization within the bit. This sense of rotation can be detected with the use of a polarizing crystal.

Current work is directed toward: (1) preparation of thin ferromagnetic films suitable for magneto-optic information storage, and (2) an experimental investigation of the Curie-point-switching process. It is in the first area that most progress has been made within the past few months.

The procedure for preparing thin ferromagnetic films of manganese bismuthide having an easy axis of magnetization perpendicular to the plane of the film, and a large Faraday effect has been more clearly defined. In this procedure, manganese, bismuth, and silicon oxide are sequentially deposited onto a freshly cleaved mica substrate, the silicon-oxide layer serving as a protective coating. The layered structure is then annealed in vacuum to yield the ferromagnetic compound manganese bismuthide. The problem has been to find the annealing procedure yielding optimum MnBi films. This problem has been overcome with the use of a vacuum furnace which allows observation of the films magnified by a factor of 200 through a crossed set of polarizers during annealing. The

temperature-versus-time profile necessary to yield optimum samples and the many interesting phenomena observed during annealing will be described in a letter to be submitted for publication in the open literature.

A literature search has yielded several other ferromagnetic compounds which might be attractive as media for magneto-optic information storage. One of these compounds is chromium telluride. The uniaxial anisotropy of this material is sufficiently high so that it can conceivably be prepared with an easy axis of magnetization perpendicular to the plane of the film. If chromium telluride has a Faraday effect comparable to that for manganese bismuthide (unfortunately, the magneto-optic properties of this material are not known), it might prove to be a more suitable material for magneto-optic information storage. The Curie temperature of chromium telluride is lower than that of manganese bismuthide (60°C as compared to 345°C), while the decomposition temperature of the former material is much higher than that of the latter (approximately 1000°C as compared to 435°C). A lower Curie point and a higher decomposition temperature relax the restrictions imposed on the power of the optical beam used to carry out the Curie-point-switching process. Equipment has been assembled for an attempt to grow thin ferromagnetic single-crystal films on mica substrates by epitaxy from the vapor phase.

C. Apparent Work Function of Cavity Emitters,

K. Shimada

1. Introduction

Unusually low apparent work functions of a 19-cavity emitter were reported in SPS 37-46, Vol. IV. According to these results, work functions were 0.4 eV lower than those expected from the Rasor-Warner theory and, therefore, the emission-current density of the 19-cavity emitter was nearly 10 times larger than that of ordinary flat emitters. This investigation was extended to a 7-cavity emitter in order to study the effect of cavity configuration on the emission property. However, the expected difference in electron emission between the 19- and 7-cavity emitters was not observed. In this article, the results obtained from the 7-cavity emitter are presented.

2. Experimental Results

The 7-cavity emitter was made of tantalum and had 7 shallow cylindrical cavities (depth = 0.0407 cm, diam = 0.396 cm). The configuration was similar to that of the 19-cavity emitter, with approximately half of the projected

Table 2. Emitter dimensions

Parameter	Seven-cavity emitter	Nineteen-cavity emitter
Cavity diameter, cm	0.396	0.236
Cavity depth, cm	0.0407	0.0407
Bottom area per cavity, cm ²	0.123	0.0437
Total bottom area A_B , cm ²	0.862	0.831
Side wall area per cavity, cm ²	0.0506	0.032
Total side wall area A_S , cm ²	0.354	0.573
Projected area A_P , cm ²	2.00	2.00
Total emitter area $A_T = A_P + A_S$, cm ²	2.354	2.573
A_T/A_P , %	117.7	128.6
A_B/A_P , %	43.1	41.5
A_S/A_P , %	17.7	28.6

emitter area of 2 cm² occupied by the bottoms of the cavities. The interelectrode spacing at an emitter temperature $T_E = 1400^\circ\text{C}$ and a collector temperature $T_C = 400^\circ\text{C}$ is 0.005 cm, and therefore the bottoms of the cavities are 0.0457 cm from the molybdenum collector. Table 2 shows pertinent dimensions of the 7- and 19-cavity emitters. Both emitters were mechanically ground to remove excess burrs from the rims of the cylindrical cavities that resulted from the drilling operation, but the rectangular edges were intentionally preserved to maintain well-defined side wall areas. The volt-ampere characteristics of the diode with the 7-cavity emitter were obtained with an X-Y recorder as the voltage across the diode was swept from -3 to +5 V. Those volt-ampere curves obtained from the unignited mode of the diode operation were used in the subsequent analysis. The saturation currents were then determined. The results are shown in Fig. 3 as a family of S-curves. The emitter temperature ranged between 1200 and 2100°K and the cesium reservoir temperature between 393 and 453°K. The cesium temperature was kept above 393°K since the error of saturation-current measurements at lower temperatures became considerable (30%) because of the smallness of the current and the lack of clean saturation.

At a cesium reservoir temperature $T_{Cs} = 453^\circ\text{K}$, the electron-neutral mean-free-path is approximately twice the interelectrode distance at the location of a cavity. Therefore, the diode was operating in a collision-less regime for cesium temperatures below 453°K. In fact, for T_{Cs} larger than 453°K, the diode ignited at relatively small voltages, and the saturation region of the volt-ampere curves became obscured. The four S-curves shown in Fig. 3 appear to converge along a straight line representing the vacuum-emission current from an emitter with the work function $\phi_0 = 4.3$ eV, which can be taken as the uncesiated (vacuum) work function of the 7-cavity emitter.

Table 2. Weld schedules for lead-interconnect cross-wire welds

Lead material	Interconnect material	Electrode force, lb	Energy, W/s	Average pull strength, lb
Au (plated) Cu 0.025-in.	10 Ag-90 Pd	7.5	14	12.1
	20 Ag-80 Pd	7.5	14	13.6
	10 Cu-90 Pd	7.5	14	12.1
	Alloy 90	8	20	12.8
	Nickel	7.5	15	14.4
Au (plated) Dumet 0.020-in.	10 Ag-90 Pd	7	4	7.8
	20 Ag-80 Pd	7.5	5	10.6
	10 Cu-90 Pd	7	5	11.1
	Alloy 90	7	7.5	10.2
	Nickel	7	6	13.6
Au (plated) Kovar 0.018-in.	10 Ag-90 Pd	7	3	7.2
	20 Ag-80 Pd	6	3	8.0
	10 Cu-90 Pd	7	3	10.0
	Alloy 90	6.5	4	8.7
	Nickel	8	3.5	16.8
Au (plated) Ni 0.025-in.	10 Ag-90 Pd	6.5	6	18.6
	20 Ag-80 Pd	6.5	8	17.8
Alloy 99 0.020-in.	10 Ag-90 Pd	7	5	13.4
	20 Ag-80 Pd	7	7	15.5

schedule for each interconnect material welded to each lead material appears in Table 2.

Evaluation and comparison of the three selected alloys to Alloy 90 and nickel ribbon, after mechanical testing and metallurgical examination, yield the following results: Splice-type lap welds of the interconnect material can be made, but the strength is much lower than that of the parent material. Alloy 90 and nickel ribbon show a smaller strength decrease when lap-welded. Cross-wire welds of interconnect material to leads show strengths less than those of Alloy 90 and nickel ribbon-welded to the same materials. Welds exhibiting higher strengths were obtained; however, these welds are accompanied by high setdown values of 50% or more and excessive expulsion and spitting of material. Decreasing the electrode force to eliminate high setdown values and decreasing the energy to eliminate excessive expulsion can only be done at a sacrifice to average pull strength.

It is presently felt that the Ag-Pd and Cu-Pd alloys selected for evaluation as nonmagnetic interconnect materials are not superior to the presently used Alloy 90 material, and that they do not compare favorably with nickel interconnect ribbon. Therefore, further evaluation of these materials has been discontinued. Further investigations into other binary alloy systems and possibly some ternary systems are being examined.

B. Planetary Entry Heat Shields, T. F. Moran

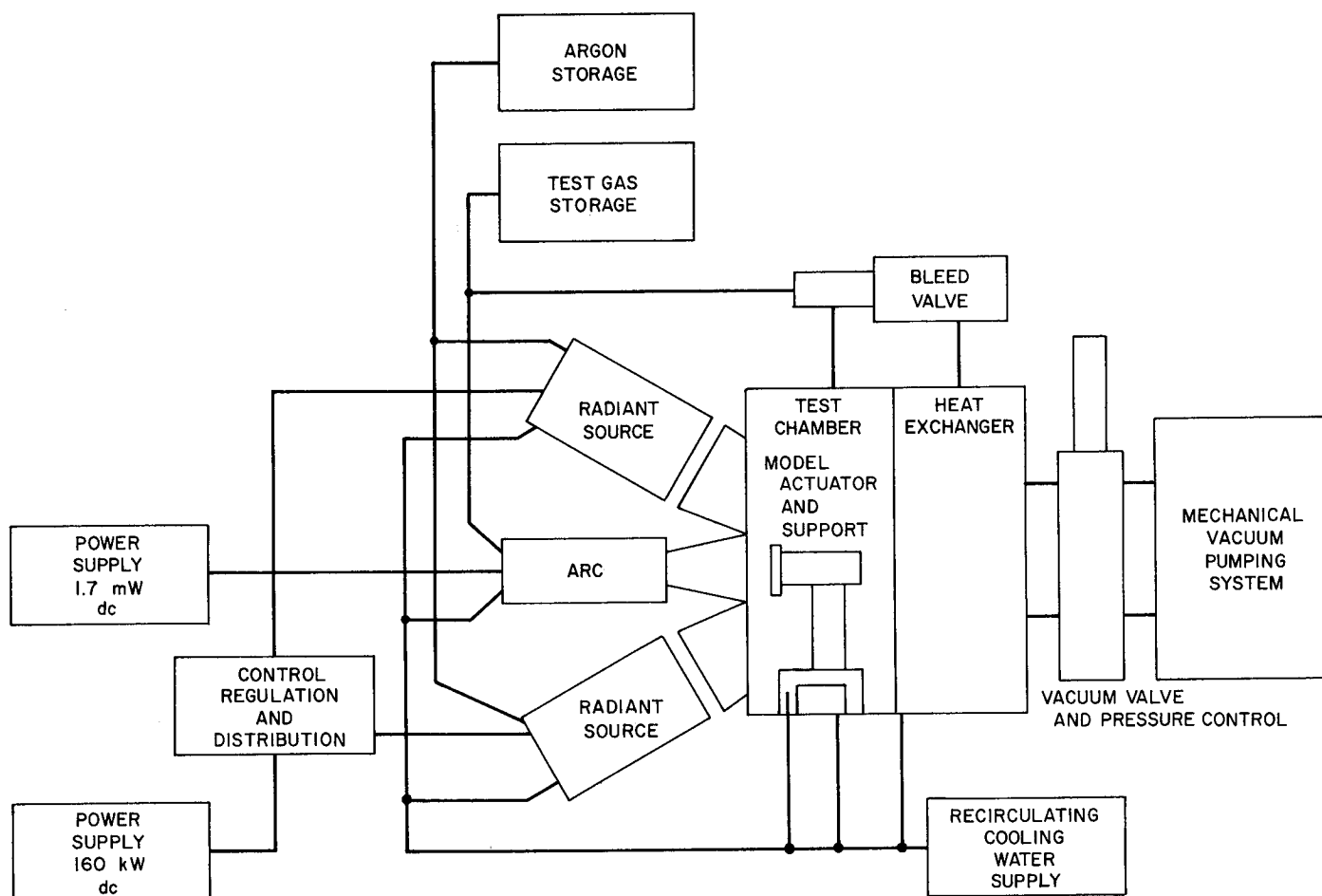
The success of future planetary entry missions depends upon the ability of the landed package to survive the severe heating encountered during atmospheric penetration. The selection of a heat shield material for assuring this survival requires a firm engineering understanding of the candidate material's response to anticipated thermal environments. One of the principal sources of data on this response comes from ground-based simulation facilities. In order to correctly evaluate and apply the plethora of data generated by such external facilities, experience in their use is imperative.

For the past 2 yr JPL has been involved in the planning and design necessary for the reestablishment of an earlier plasma generator facility (Ref. 1) to provide JPL with a limited capability in this area. The purpose of the facility is not to eliminate the need for external testing in this area, but to provide JPL with a knowledge of the limitations and fallacies of the testing technique. This will allow critical JPL recommendations or decisions for specific planetary entry studies or projects to be based on direct experience.

In order that testing, representative of Martian atmospheric entry and, to a more limited degree, Venusian entry, may be accomplished, it was necessary to include

Upon completion, a limited amount of ablative material evaluation and external data checking will be able to be accomplished within the facility's parametric envelope (pressure, enthalpy, specimen size, heating rate, test duration, etc.). It is also possible that some research into problem areas for which larger facilities do not have the time available may be done. As initial construction nears completion, it appears to be an appropriate time to describe details of some of the progress to date.

The vortex stabilized plasma generator unit used in the facility has been described earlier (Ref. 1). A working gas is tangentially injected into the generator, enters an excited state upon passing through the spinning arc column, is brought to equilibrium in a plenum chamber and exhausts through a supersonic nozzle. The particular unit presently set up is rated for maximum input power levels of 600 kW with transfer efficiencies of less than 60% depending on chamber pressure and mass flow rates.



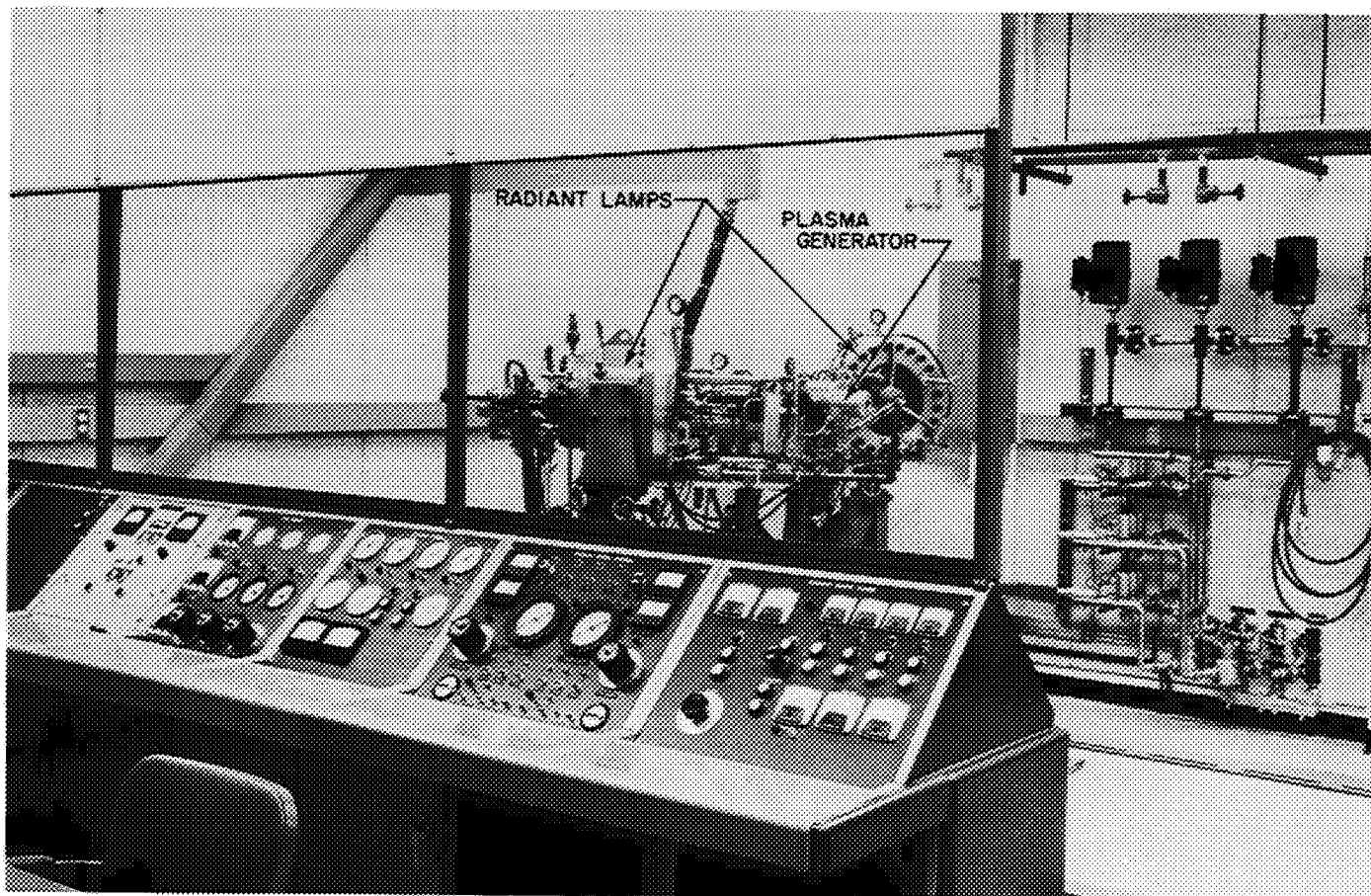


Fig. 2. Extraterrestrial atmospheric entry test facility

Input power for the plasma generator is supplied by a 1.7-MW dc silicon-controlled rectifier supply designed and built by Cal-Power Corp. The supply is designed to use continuous power variation to approximate the actual pulsed heating profiles of entry. Each of the four modules of the supply (shown in Fig. 3) has a rating of 1250 A and 350 V. Series and parallel combinations will allow up to 5000 A or 1400 V output.

Complementing the convective plasma generator are two radiant lamps designated Avco Model RAS-3A. The lamps may be run and controlled either independently or simultaneously with the plasma generator. Each lamp consists of a plasma radiation source and an optical system for concentrating this energy on a remote target area. A lamp is shown schematically in Fig. 4. Cold gaseous argon is delivered to the lamp at pressures up to 20 atm. As the argon flows through the arc between two specially impregnated electrodes it is converted into a high-energy plasma which emits radiation with an intensity directly

proportional to the current and the square root of the pressure. The radiation is collected by two reflectors, transmitted through a quartz window and focused externally with a decentered torroidal biconvex concentrating lens. The reflectors are constructed of polished aluminum and are rhodium plated to help prevent fogging. The lamps are mounted on indexing heads to provide variable positioning. Input power is supplied by a bank of moving-coil rectifiers. When connected in series, they are rated at 160 kW dc and will deliver stable currents up to 1000 A. An additional moving-coil rectifier is used to supply low-voltage power to individual field coils on the lamps for magnetically rotating the arc.

A high-voltage, high-frequency oscillator supplies the breakdown voltage to start the lamps. A pneumatically operated filter and douser are provided to control the output radiation to the target area. The douser is also instrumented as an absolute calorimeter, thus providing a relative flux calibration.

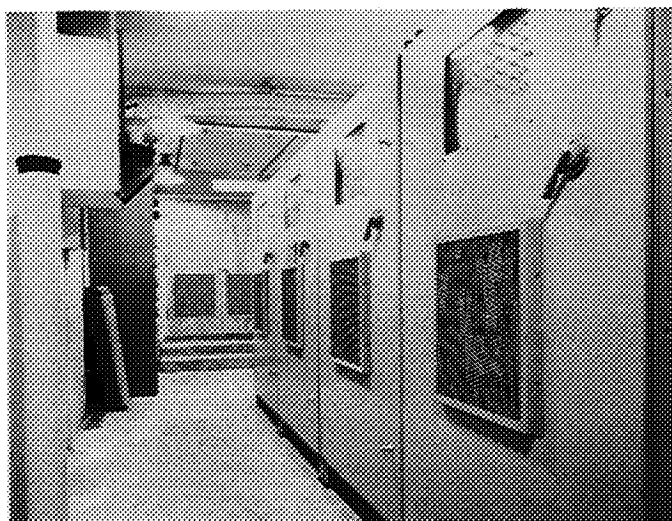


Fig. 3. Silicon-controlled rectifier power supplies

Each lamp can provide a variable flux output from 0 to 250 Btu/ft²/s over an 0.750-in.-diam target area. Maximum output is obtained at input levels of approximately 800 A and 150 psig.

The flux variation across the 3/4-in.-diam target area is reported by the manufacturer to be less than $\pm 10\%$. Time stability is such that test position flux will not vary more than $\pm 10\%$ in 10 min of operation. Flux reproducibility is within $\pm 10\%$ over the entire flux range.

A 280 gal/min recirculating water supply is provided for cooling the lamps, plasma generator, and peripheral equipment. A booster pump can supply up to 180 gal/min of this water at pressures up to 500 psig. The remaining 100 gal/min can be supplied at 200 psig. Working gas is supplied from bottle storage banks. Both coolant and gas flow rates and temperatures can be continuously monitored and recorded.

Design of a vacuum test chamber which will allow a model to be exposed to both radiative and convective heating under controlled pressure conditions has been

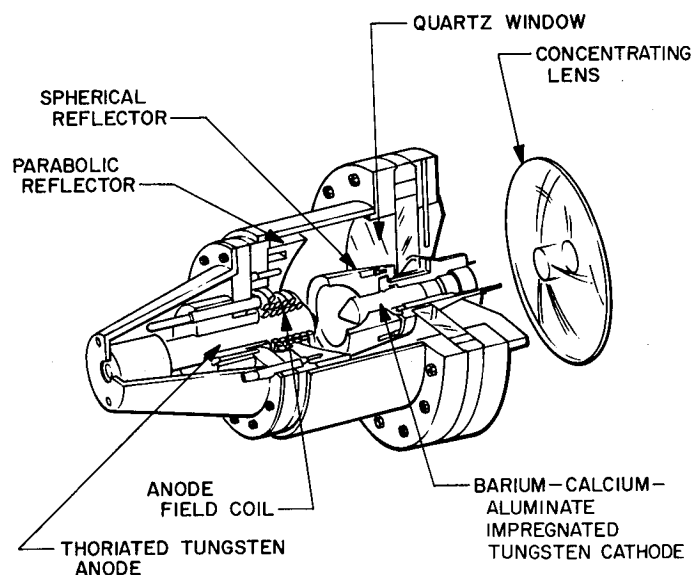


Fig. 4. Sketch of radiant arc lamp

completed. The cylindrical tank volume is approximately 115 ft³ and includes ports for radiant and convective energy access, test stream viewing, sample insertion, vacuum connection, nitrogen bleed, and instrumentation feed-throughs. The water-cooled tank contains a cone shaped tubular heat exchanger downstream from a flow deflector to provide ambient temperature gases at the vacuum pump inlet.

Future additions include a model insertion system to allow rapid insertion and extraction of test specimens from the plasma stream and accurate low velocity axial and radial scanning of the test stream. A mechanical vacuum pumping system has been tentatively specified for use in evacuating the chamber and for fine control of the pressure under dynamic conditions.

Reference

1. Nagler, R., "Endothermal Materials," Research Summary 36-4, Vol. II, pp. 25-28, Jet Propulsion Laboratory, Pasadena, Calif., Sept. 1, 1960.

IX. Electronics Parts Engineering

ENGINEERING MECHANICS DIVISION

A. Accelerated Testing Concepts, Methodology, and Models: A Literature Review, E. Klippenstein

Basic problems in assessing the reliability of electronic parts are the long time and large-scale life tests which are required to demonstrate low failure rates. A solution to these problems is accelerated life tests. The usefulness of accelerated testing, however, is limited to the extent that performance of parts at accelerated conditions can be validly related to performance at normal use conditions.

In order to obtain a rigorous and critical review of what has been done in the area of accelerated testing of electronic parts, a contract was let in February 1967, to perform a review of the literature. The review period was for 1 yr and to cover a significant portion of all available literature.

At the present time, 82 articles and reports have been critically reviewed, and a 6-mo summary report has been written. A total of 247 articles and reports have been identified as pertinent. The review is continuing to cover a significant portion of the remaining literature. A final summary report will be written in February 1968.

Based on the reviews and the 6-mo summary report, this article will be devoted to concepts, methods of pro-

gramming severity level, and models relating some parameter (e.g., failure rate) to severity level.

1. Concepts

a. Conceptual model. The idea of a conceptual model is to describe characteristics of importance and ignore the remainder. Since an exact treatment would be very complicated, we must state our assumptions and then operate on those assumptions with mathematics and reason. The assumptions, together with the current results of the analysis, are our model.

b. Simple stress-strength model for failure. There is a value of stress called the strength such that there is failure if, and only if, the instantaneous stress exceeds the strength. Stated another way: There exists a scaler S which can only depend reversibly on the environment of the part, and a value S^* of that scaler such that the part fails if and only if $S > S^*$; S^* is the strength. Values of $S < S^*$ do no damage.

c. Simple damage-endurance model for failure. The application of a damager causes cumulative damage in some way and some of the endurance of the part has been consumed even if failure does not occur. When the damager has been removed, the damage is not undone. Stated another way: There exists a scaler D which depends on

the set of damagers and on the behavior over time and a value E of that scaler called the endurance such that failure occurs if, and only if, $D > E$. For $D < E$, the amount of endurance remaining is $E - D$. Note that this model is usually assumed when dealing with life tests and applying accelerated damagers.

d. Hazard rate and damage. The endurance of an individual part is generally a random variable. For a constant hazard rate process with constant environmental severity, damage is done as time goes by; therefore, a used part is not as good as a new one. But since we do not know the life of an individual part, the distribution of part lives can be such that any part known to be good is as likely to last as long as any other part known to be good, whether new or not. For a decreasing hazard rate process, even though a part is being degraded, as long as it has not failed, it is more likely to last longer than a part that has not yet been used.

e. True acceleration. Acceleration is true if, and only if, the system passes reasonably through equivalent states in the same order as it would at usual conditions. The word "reasonably" is necessary because as engineers, if things are close enough for the purpose, it satisfies the requirements of that situation. Stated another way: Let $g(t)$ be the state of the system under usual conditions and $G(t)$ be the equivalent state under accelerated conditions, then there is true acceleration if, and only if:

- (1) $G(t) = g(\tau[t])$
- (2) $\tau(t)$ is a monotonically increasing function
- (3) $G(0) = g(0)$
- (4) $\tau(0) = 0$

The acceleration factor A is defined as $A(t) = \tau(t)/t$

2. Methods of Programming an Accelerated Test

a. Constant-stress test. This is the type of test normally used in rated life tests wherein the stress level remains constant throughout the life of the items on the test. In accelerated testing programs, it is customary to run tests at several severity levels and to plot a curve showing some measure of goodness versus the measure of stress. The measure of goodness may be failure rate, time to failure, etc.

b. Step-stress test. In this type of test program, the stress level is increased in increments at uniform time intervals. It is customary to run tests with several groups, varying the time interval for each group. It is convenient to dis-

tinguish between large, medium, and small steps. In large step tests, it is assumed that the cumulative damage is negligible up to the last step before failure. In medium step tests, the cumulative damage at previous steps must be taken into account, but the steps are not small enough that the severity level is continuously increasing. In small step tests, the steps are small enough so that one can assume with negligible error that the severity level is steadily increasing. Tests using large steps are analyzed as if they were constant-stress tests being run at the severity level of the failure step. Medium and small step-stress tests are analyzed accounting for cumulative damage; for medium steps by summation at each prior step; and for small steps or progressive tests by integration of previous cumulative damage. In order to relate the results of the various tests, one needs some theory of cumulative damage. The one usually chosen is the simple linear theory.

c. Progressive stress test. In this type of test, the stress level is increased at a fixed rate until failure occurs. Several groups of parts are tested wherein each group is subjected to a different rate of increase in stress. The results are analyzed accounting for cumulative damage.

d. Other programs. A simple modification to the step-stress and progressive stress tests is to start the severity level above zero. The term "probe testing" is used sometimes but this is a special case of step or progress stress testing where the stress is a vector of several dimensions. Some test programs change severity level only once with the first severity level being high and the second low. This type of test is used to investigate cumulative damage theories.

3. Conceptual Models for Accelerated Stress

a. Arrhenius equation. There are situations where rates of reactions are dependent on temperature. It appears that temperature is one of the most important damagers that we have. A classic example for temperature dependence of specific reaction rates is the following equation credited to Arrhenius:

$$rr = A \exp(-E/kT)$$

where

rr = reaction rate

A = a constant (also called frequency factor)

E = activation energy (eV/molecule)

k = Boltzmann's constant

T = temperature

In using this conceptual model to relate some parameter (measure of goodness) to the test temperature, one usually plots the log (results) versus $1/kT$ (or against $1/T$) and hopes to get a straight line which is desirable for extrapolation.

b. Eyring equation. Another equation relating reaction rate to temperature is the following equation credited to Eyring:

$$rr = \frac{\kappa kT}{h} \exp\left(\frac{-\Delta G}{kT}\right)$$

where

κ = transmission coefficient

h = Planck's constant

G = Gibb's free energy

Since $\Delta G = \Delta H - T\Delta S$ where H and S are enthalpy and entropy, the equation can be put in the following form:

$$rr = \frac{\kappa kT}{h} \exp\left(\frac{\Delta S}{k}\right) \exp\left(\frac{-\Delta H}{kT}\right)$$

The ΔH is closely associated with activation energy. The term $\exp(\Delta S/k)$, however, gives trouble. A potential energy surface can be introduced and if there are not too many dimensions and the system is extremely simple, then this surface can be obtained from quantum mechanical considerations.

The Eyring equation has been touted as a relationship of fundamental quantities and therefore should be used as the conceptual model for ageing. Actually, electronic components are complex engineering systems from the point of view of theoretical chemistry/physics and for practical purposes the Eyring equation will offer little, if anything, over the Arrhenius equation.

c. Voltage law for capacitors. The voltage law, or more commonly called the "fifth power rule," is sometimes written as follows:

$$\frac{L_1}{L_2} = \frac{R_1}{R_2} \left(\frac{V_2}{V_1}\right)^n \times 2^{[(T_2-T_1)/K]}$$

where

L = time to given percent failure

R = percent failures on which L is based

V = dc voltage

T = temperature, °C

n = power law exponent

K = temperature constant

The rules of thumb are that $n = 5$ and $K = 10^\circ\text{C}$, which says that life varies as the fifth power of the dc voltage and doubles for every 10°C decrease in temperature. These rules are largely empirical and are sometimes useful. The literature gives values of n from 3 to 8 and suggests that n may be a function of the voltage. The value of K may be 5 to 30°C and may also be a function of the temperature.

4. Future Plans

The plans for the remainder of this work effort are: to continue reviewing the literature; to work out some specific examples of analysis assuming a constant hazard rate process and the Arrhenius model and, in particular, calculating the uncertainties; and to obtain a comprehensive final report. There is strong momentum toward the extended use of the "physics of failure" approach and the use of small-quantity short-time accelerated tests. With a comprehensive report on accelerated testing, it is hoped that the information will help in avoiding the pitfalls which are possible and provide an awareness of the uncertainties and risks, recognizing them for what they are. The usefulness of accelerated tests probably dates back to the caveman when he banged his weapon on a rock to see if it would stand the impact. Accelerated tests are still useful in measuring strength, estimating endurance, and predicting life, etc. The problems appear to be the credibility of predictions which are based on assumed models and where uncertainties are not calculated or are completely unknown.

X. Solid Propellant Engineering

PROPULSION DIVISION

A. Applications Technology Satellite Motor

Development, R. G. Anderson and R. A. Grippi, Jr.

1. Introduction

In January 1963 the Jet Propulsion Laboratory initiated a development program to provide a solid-propellant apogee motor for a second-generation *Syncom* satellite. This program, under the management of the Goddard Space Flight Center, was designated *Advanced Syncom*. It was to result in a spin-stabilized, active repeater communications satellite weighing about 750 lb, operating at synchronous altitude (22,300 mi) to handle voice communications, teletype, and monochrome and color television signals.

In January 1964 the *Advanced Syncom* communication program was redirected to include a number of experimental instruments, in addition to the original communication instruments. This expanded program is the *Applications Technology Satellite* program and will result in a general-purpose satellite capable of operation at synchronous altitude with experimental instruments in the areas of meteorology, communications, radiation, navigation, gravity gradient stabilization, and various engineering experiments. For those satellites to be placed in synchronous orbit, JPL will provide a solid-propellant

rocket motor to provide final required velocity increment at the apogee of the elliptical transfer orbit. This rocket motor is designated the JPL SR-28-1 (steel chamber) or JPL SR-28-3 (titanium chamber). At present, only the JPL SR-28-3 unit is intended for flight use.

Previous reports of progress on the development of this motor have been published in SPS 37-20 to 37-33, Vol. V, SPS 37-34 to 37-45, Vol. IV, and SPS 37-47, Vol. III.

2. Program Status Summary

The motor development program calls for static firing of four heavywall motors and 26 flightweight motors, including two with flight design titanium chambers, prior to conducting an eight-motor qualification program. To date, the four heavywall motors plus 25 flightweight motors have been static-fired, four of which were under simulated high-altitude conditions at AEDC. All of the flightweight motors tested to date have been with type 410 chromium steel chambers with the exception of Dev. G-8T, G-9T, E-3T, and Q-9T, which used titanium chambers.

The ATS apogee motor qualification phase was conducted at AEDC during July and August of 1966. The results are reported in SPS 37-41, Vol. IV, pp. 91 to 95.

The first of five *Applications Technology Satellites*, ATS-B, was successfully launched during December 1966. The satellite is presently on station over the Pacific Ocean and functioning as planned. The second synchronous satellite, ATS-C, was launched November 5, 1967. Approximately 16 h after launch, the JPL apogee motor was fired, transferring the satellite from its highly elliptical transfer orbit into a near-synchronous equatorial orbit. The ATS-C satellite is presently on station over the Atlantic Ocean, with all experiments functioning. Figure 1 shows a model of *Applications Technology Satellite* and JPL apogee motor.

The remaining JPL ATS effort of any scope will be the loading of three to four flight apogee units during the early part of 1968. These apogee units are designated for the ATS-D (June 1968) and ATS-E (April-May 1969) launches. JPL engineering launch preparation and support will be provided for the remaining two launches.

3. Apogee Motor Storage Phase

Three ATS apogee motors were assigned to the storage phase of the motor development program. The primary objective of the storage phase was to demonstrate that the apogee motor is acceptable for flight after extended ambient storage. Initially, it was decided to store one motor of 12 mo and two units for 24 mo. These storage periods were based upon preliminary flight loading and launch dates. However, a change in two of four launch

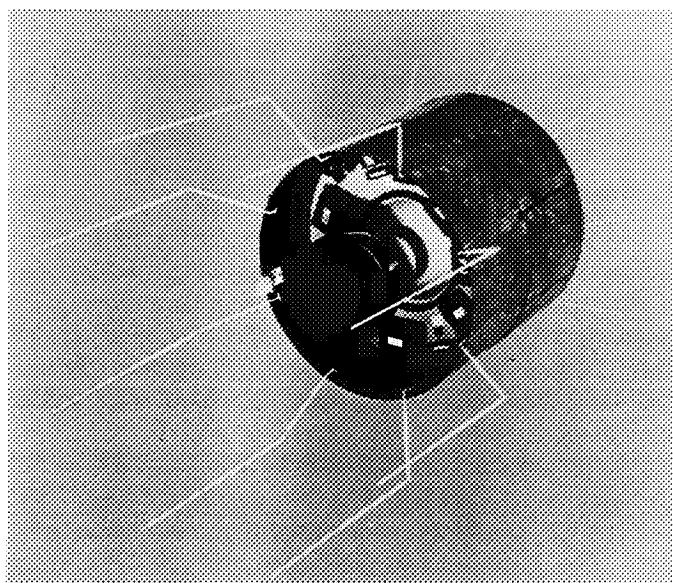


Fig. 1. *Applications Technology Satellite* and JPL apogee motor

dates and subsequent rescheduling of the flight loadings resulted in modified storage periods. The first unit was stored for 16 mo, second unit 20 mo, and the third unit 24 mo.

The three units, designated development codes F-1, F-2, and F-3, were processed and loaded to flight standard procedures in August and September 1965. All hardware components (including propellant) were flight design, with the exception of the chamber material and insulation thickness. In place of the flight-type 6Al-4V titanium chamber, type 410 chromium steel cases were assigned to the storage units. During this time period the titanium chambers were in process of manufacture. The insulation configuration, compared to flight design, lacks one layer of 0.030-in. material¹ in each dome section. The difference between the flight and storage insulation configuration was a result of changes initiated after the first titanium chamber firing. The additional layer of insulation was added to reduce the chamber temperature during motor operation and improve postfire motor balance. However, either deviation from the flight configuration should not affect the storage characteristics or performance of the motor.

Each motor was subjected to the identical processing operations. This consisted of the following major items:

- (1) The use of flight-quality hardware components.
- (2) Preloading weight determinations.
- (3) Preloading motor assembly alignment.
- (4) Preloading static and dynamic balancing.
- (5) Propellant loading to flight finalized procedures.
- (6) Visual and radiographic inspection of propellant.
- (7) Motor assembly alignment.
- (8) Motor assembly weight.
- (9) Center of gravity and moment-of-inertia determinations.
- (10) Loaded motor imbalance determination.

Upon completion of motor processing in November 1965, the units were placed in ambient storage. However, after the initial periodic inspection during March 1966, the units were stored at either 60 or 80°F for the duration of their storage period.

¹Items 17 and 18 on JPL drawing J390 1797A.

Each motor was removed from storage at designated intervals, as shown on Table 1, and subjected to the following sequential inspections:

- (1) Moisture condition of container.
- (2) Motor assembly weight.
- (3) Motor assembly leak check.
- (4) Motor assembly alignment.

- (5) Port alignment.
- (6) Port diameter measurement.
- (7) Propellant, insulation, chamber and nozzle visual inspection.
- (8) Component (nozzle and chamber) weight.
- (9) Motor assembly alignment.
- (10) Motor assembly leak check.

Table 1. ATS apogee motor. Storage phase (periodic inspection schedule)

Inspection	F-1	F-2	F-3
Cast	9/23/65	9/30/65	9/17/65
1st inspection	3/30/66	3/31/66	4/5/66
2nd inspection	8/8/66	8/10/66	8/15/66
3rd inspection	10/7/66	11/8/66	11/10/66
4th inspection	—	3/24/67	3/17/67
5th inspection	—	7/10/67	—
Final inspection	1/24/67	9/21/67	5/3/67
Static test	1/31/67	9/29/67	5/10/67
Storage period	16 mo	24 mo	20 mo

After each inspection the motors were packaged, purged with dry nitrogen, and returned to storage. Table 2 summarizes the results of these inspections. This table includes data taken after the units were loaded in September 1965 and data recorded just prior to static testing. All prefire data indicated that the apogee units could be successfully stored for extended periods.

As shown on Table 2, all data appear to be nominal, and changes which occurred during the storage period are insignificant. For example, changes in the loaded chamber weight can be attributed to insignificant scale inaccuracies at the initial weighing. Since the prestorage

Table 2. ATS apogee motor. Storage phase (weight and measurement data prestorage and poststorage)

Item	F-1		F-2		F-3	
	Pre	Post	Pre	Post	Pre	Post
Loaded chamber wt, lb (including handling ring)	868.3	868.7	870.8	870.6	870.4	870.0
Nozzle wt, lb	39.08	38.90	39.32	39.12	38.95	38.75
Visual inspection	OK	^a	^b	^b	OK	^c
Thrust misalignment, in./in.	0.00012	0.00012	0.00008	0.00013	0.00006	0.0003
Exit cone TIR, in.	0.004	0.013	0.009	0.008	0.004	0.010
Port TIR, in. ^d						
Station 4	0.016	0.015	0.009	0.009	0.006	0.005
Station 14	0.017	0.016	0.010	0.011	0.007	0.007
Station 24	0.014	0.011	0.007	0.009	0.008	0.009
Center of gravity, in.	11.49	—	11.45	—	11.46	—
Imbalance (after propellant loading)						
Dynamic, lb-in. ²	36.8	—	15.4	—	30.8	—
Static, lb-in.	3.5	—	2.7	—	3.0	—
Port diameter, in. ^e						
Station 4	10.147	10.192	10.159	10.180	10.122	10.167
Station 14	10.143	10.168	10.149	10.161	10.127	10.143
Station 24	10.139	10.140	10.140	10.138	10.122	10.131

^aPropellant-to-insulation separation at head end extending 1 in. in depth. Discrepancy first observed during 3/30/66 inspection.
^bInsulation-to-chamber separation at aft end extending 0.3 in. in depth. Discrepancy first observed after propellant loading.
^cPropellant-to-insulation separation at head end extending 1 in. in depth. Discrepancy first observed during 11/10/66 inspection.
^dStation number represents distance in inches from nozzle boss surface.
^ePort diameter dimensions taken at a propellant bulk temperature of 60°F.

weighing of these units, the weighing procedures have been changed to improve upon the standard 0.1% scale accuracy. The benefits derived from these changes became evident during the periodic inspection weighings. Weights recorded during these inspections and the final poststorage weights were consistent within the scale readability, ± 0.1 lb. Therefore, the loaded chamber weight data has demonstrated that the apogee motor propellant configuration can be stored for extended periods without incurring any weight gain or loss.

The nozzle assembly weight data, as expected, indicates a weight loss. This change can be explained by the fact that the phenolic resin system in the carbon and silica ablative material will outgas after the composite is cured. This occurs more readily when an ablative material has a machined surface with respect to an as-cured surface. All surfaces of the nozzle have a machined surface. Further detailed investigation of the weight loss in the storage nozzles and other ATS nozzles have shown that the rate of off-gassing decreases asymptotically.

Visual inspection of the propellant to insulation and insulation to chamber interfaces has indicated satisfactory storage, except for three insignificant separations. The separation visible on F-1 was located at the forward opening of the motor and was first evident at the initial periodic inspection (March 30, 1966). However, the defect was minor, i.e., approximately 1 in. in depth, and the motor was successfully fired without repairing the defect. Motor F-3 also incurred a propellant-to-insulation separation at the forward end. The discrepancy was first evident during the November 10, 1966 inspection. Again, the defect was minor, extending 1 in. in depth, and the motor was successfully fired without repairing the defect. After propellant loading a portion of the insulation to chamber bond failed on motor F-2. This defect extended 0.3 in. into the interface and approximately 120 deg around the periphery of the aft opening. Prior to static test, the defect was repaired and the motor fired without incident in this area.

The slight changes in the motor port diameter, shown on Table 2, can be attributed to the typical viscoelastic properties of polyurethane propellant. As expected, the port diameter increased (propellant shrinkage) and the propellant slumped downward toward the igniter opening of the motor.

In January 1967 the first storage unit, Dev. F-1, was removed from storage, inspected, and subjected to environmental testing. The test environments consisted of temperature cycle, booster vibration, and booster acceler-

ation. After each category of environmental testing, the motor underwent extensive inspections. This included all inspections previously mentioned and radiographic inspection of the nozzle and propellant. The igniter basket was also visually inspected for anomalies. No discrepancies were noted on motor F-1 as a result of the long-term storage and subsequent environmental tests.

After the environmental tests the motors were inspected, reassembled (with the original O-ring seal), and conditioned to 10°F. Table 3 summarizes the prefire motor configuration. Then the units were removed from the 10°F environment, positioned on the JPL-ETS vertical spin stand, and successfully static-fired at 150 rpm. The first unit (Dev. F-1) was subjected to 16 mo of ambient storage prior to testing on January 31, 1967; the second unit (Dev. F-3) was tested on May 10, 1967, after 20 mo of ambient storage; and the third unit (Dev. F-2) was subjected to 24 mo of storage and was tested on September 29, 1967. The three units operated normally during their 45-s operation. Postfire inspections of the three spent motors indicated nominal operations. Table 4 summarizes the three motors' performance parameters and lists other pertinent data related to the units.

Table 3. ATS apogee motor. Storage phase (identification and final weight summary)

Identification	Part No.	Serial No.		
		F-1	F-2	F-3
Assembly	—	F-1	F-2	F-3
Chamber	J3901513E	P-32	P-34	P-23
Insulation	J3901797 N/C	P-32	P-34	P-23
Nozzle	J3901659D	F-26	F-27	F-28
Igniter basket	J3901802 N/C	SYC-257	SYC-258	SYC-259
Propellant	JPL540K	SY-273	SY-274	SY-275
Squib	SDI 101120(F-1) SDI 100728 (F-2, F-3)	56	34	35
Final weight data, lb		F-1	F-2	F-3
Chamber (including balance wt)		41.11	40.40	40.63
Insulation		10.57	10.98	11.37
TDI (insulation rinse)		0.33	0.40	0.38
Nozzle (including balance wt)		38.90	39.12	38.75
Igniter basket		1.01	1.01	1.01
Miscellaneous (nozzle bolts, washers, and O-ring)		0.35	0.35	0.35
Propellant		764.4	766.3	764.7
Motor assembly weight (including igniter basket)		856.7	858.6	857.2
Total motor inert weight (including igniter basket)		92.3	92.3	92.5

Table 4. ATS apogee motor. Storage phase (static test data summary)

Parameters and conditions	F-1 ^a	F-2	F-3 ^a
Test conditions			
Type of test	Atm-spin	Atm-spin	Atm-spin
Test location	JPL-ETS	JPL-ETS	JPL-ETS
Date	1/31/67	9/29/67	5/10/67
Run No.	E-781	E-829	E-792
Grain temperature	10°F	10°F	10°F
Propellant weight, lb	764.7	766.3	764.7
Pressure data			
Characteristic velocity, W*, ft/s	4896	4896	4903
Chamber pressure integral, psia/s	8992	8892	8866
Igniter peak pressure, psia (ms)	^b	1892 (15)	1607 (35)
Chamber ignition peak pressure, psia (ms)	230 ^b	225 (38)	235 (44)
Chamber starting pressure, psia (s)	103 (0.20)	99 (0.21)	98 (0.21)
Chamber run peak pressure, psia (s)	242.2 (34.3)	248.1 (32.9)	244.1 (32.2)
Time			
Ignition delay, ms	^b	16	20
Run time, s	44.96	44.53	44.54
Nozzle dimensions			
Throat diameter, in.			
Initial	4.083	4.083	4.083
Final	4.106	4.104	4.103
Average	4.095	4.093	4.093
Throat erosion (area), %	1.13	1.03	0.98
^a All pressure data approximately 1% low.			
^b Amplifier malfunction, data not available.			

As shown on Table 4, all static test performance data appear nominal, with the exception of the characteristic velocity value W^* for motors F-1 and F-3. The measured W^* values, in both cases, are approximately 1% below the average value obtained from the development and qualification units. A detailed instrumentation calibration check after the second test (Dev. F-3) revealed that the spin stand's slip rings were badly worn. This worn condition caused additional line resistance that was not taken into account during the pre- and postfire calibrations for tests F-1 and F-3. Subsequently, it was demonstrated that the worn slip rings contributed at least a 0.5% error by causing reduced motor pressure measurements. However, it was still felt that the total performance loss (1%) was created by the worn slip rings, and it was not associated with long-term storage of the apogee motor. Therefore, a new 50-channel slip ring was purchased, installed, and calibrated for use in the final storage motor (Dev. F-2) test. As indicated in Table 4 the performance (W^*) of the third motor after 24 mo of storage is nominal when compared to motors tested in the development and qualification phases. Therefore, it is felt that the worn slip rings were responsible for the total measured performance loss (1%) of motors F-1 and F-3.

Additional investigation revealed that the slip rings had been removed from the spin stand after the last development phase test, Q-9T, January 1966. Subsequently, the slip rings were inadvertently worn while in another application and unintentionally returned to the spin stand for the F-1 test. In conclusion, it is felt that the worn slip rings did not contribute any errors to motors tested in the development phase.

In summary, three flight-type apogee motors have been stored for periods of 16, 20, and 24 mo without incurring any significant changes in configuration. During the storage period these units were closely monitored and inspected at periodic intervals to observe the occurrence of any possible changes. As anticipated, no significant discrepancies occurred during the storage periods. Following the storage interval each unit was subjected to environmental tests which exceeded actual flight conditions. Then the motors were reinspected and static tested at 10°F while spinning at 150 rpm. Final performance data demonstrates that the apogee motor will perform nominally after an extended storage period.

B. Nozzle Thrust Misalignment, L. D. Strand

1. Program Status Review

An investigation is being conducted into the effects of nozzle surface irregularities and throat asymmetry on the

position of the nozzle thrust vector. Such irregularities and asymmetry can result from the delamination and/or erosion of nozzle ablative materials during rocket firing.

The experimental portion of the program consisted of cold-flow tests conducted using the Aerodynamic Facilities' auxiliary flow channel. The thrust misalignment was measured for two test nozzle systems. The steel nozzles were conical and of equal throat area and expansion ratio. The first configuration was an axisymmetric nozzle tested with and without a flow protrusion. The second nozzle was fabricated with a known throat region asymmetry. Both nozzles were instrumented with over 80 static pressure taps. An analytical approximation of the force unbalance normal to the throat nozzle axis was obtained by numerically integrating the pressure distribution over the nozzle wall. A complete description of the test facility and the test program and results was reported in SPS 37-35, Vol. IV, pp. 130 to 140.

The results of the theoretical analyses pursued to date were reported in SPS 37-37, Vol. IV, pp. 124 to 130. They included presentation of the overall side-force data as side-force axial profiles of the net side force summed over the nozzle axial distance and comparison of the measured pressure data for the asymmetric nozzle with the results of a semigraphical two-dimensional method-of-characteristics (MOC), nozzle-flow hand calculation. A combined one-dimensional isentropic, three-dimensional MOC flow-analysis method was used to obtain a qualitative picture of the flow through the asymmetric nozzle.

The summed net side-force axial profile for the asymmetric test nozzle reached a maximum value, continued to decrease as the nozzle was traversed along its axis (increasing nozzle expansion ratio ϵ), and finally leveled off at a net side force/nozzle supply pressure ratio of 0.03 (Fig. 2). The reversal in the force profile was due to changing nozzle surface pressure profiles with increasing axial distance and was determined to be attributable not solely to the particular geometry of the test nozzle.

The problem remained of attempting to extend the present experimental results to nozzle conditions of reduced asymmetry, more in line with the magnitudes of asymmetry that can occur in actual rocket nozzles (erosive increases in the nozzle throat region radii of several ten thousandths of an inch). To assist in the solution of this problem, a computer program was developed that enables the nozzle-flow characteristics for a nozzle with any degree of asymmetry to be determined. The decision was made to limit the analysis to two-dimensional nozzles, as

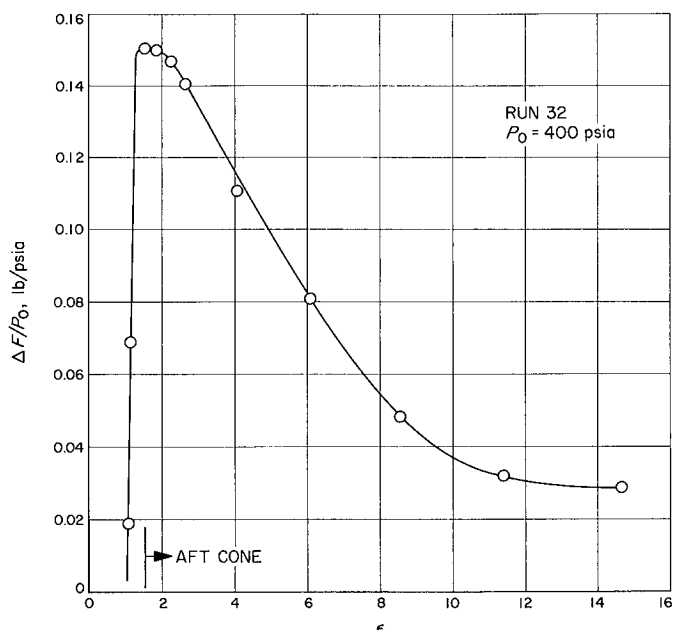


Fig. 2. Summed net side force/supply pressure ratio versus nozzle expansion ratio, asymmetric nozzle

the complexity of a three-dimensional nonsymmetric program would be prohibitive. The calculated results would attempt to be related to the three-dimensional case, using the existing experimental information.

2. Nozzle-Flow Analysis Computer Program

A Boeing Scientific Research Laboratory MOC computer program (Ref. 1) for the analysis of two-dimensional or axially symmetric, isentropic or variable entropy, nozzle flow of a perfect gas in the supersonic region was modified to analyze two-dimensional asymmetric nozzles with prescribed upper and lower wall boundaries, each boundary consisting of a circular arc of radius R and a conical section with a half-angle α . The program can use either an input initial Mach line (> 1) or calculate a uniform starting line of M prescribed points using uniform spacing between the defined boundaries. Printed outputs consist of either: (1) the coordinates, pressure ratio, Mach number, and entropy at each point where a Mach line intersects a boundary, or (2) this information, excluding pressure ratio, printed for each point of the characteristic net. A plot of pressure ratio versus axial coordinate is generated for both the upper and lower boundaries.

3. Program Calculations

The computer program was used to calculate the boundary pressure data for a family of two-dimensional

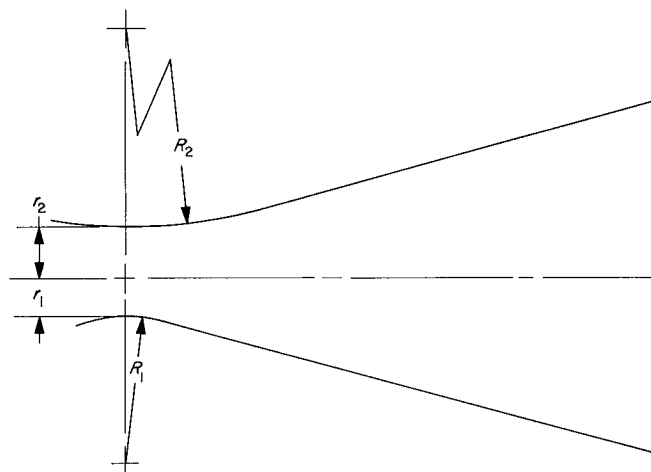


Fig. 3. Two-dimensional asymmetric nozzles

nozzles of increasing asymmetry (Fig. 3). The configuration variables were the distances from the nozzle center line to the lower and upper boundaries at the nozzle throat (r_1 and r_2), the difference being the throat asymmetry, and the radii of curvature of the circular portion of the lower and upper boundaries (R_1 and R_2). Calculations were run for the seven cases listed in Table 5. R_1 and r_1 were held fixed for all seven cases. R_2 and r_2 were increased from the Case 1 values of R_1 and r_1 (symmetric nozzle) to the Case 7 maximum values. A conical half-angle of 15 deg was used. The Case 7 configuration approached the cross-sectional wall profile of the three-dimensional asymmetric test nozzle (same r_1 , R_1 , and r_2 values; R_2 for the test nozzle equalling ∞).

Preliminary calculations were made in order to compare the nozzle pressure ratios computed using the two possible starting Mach line inputs which were calculated from, first, a modified version of the Sauer transonic flow solution and, secondly, from a uniform starting line. The

Table 5. Two-dimensional nozzle boundary variables ($r_1 = 0.526$, $R_1 = 2.04$)

Case	r_2	R_2	$r_2 - r_1$
1	0.526	2.04	0
2	0.538	2.38	0.012
3	0.550	2.72	0.024
4	0.575	3.42	0.049
5	0.600	4.13	0.074
6	0.650	5.55	0.124
7	0.705	7.11	0.179

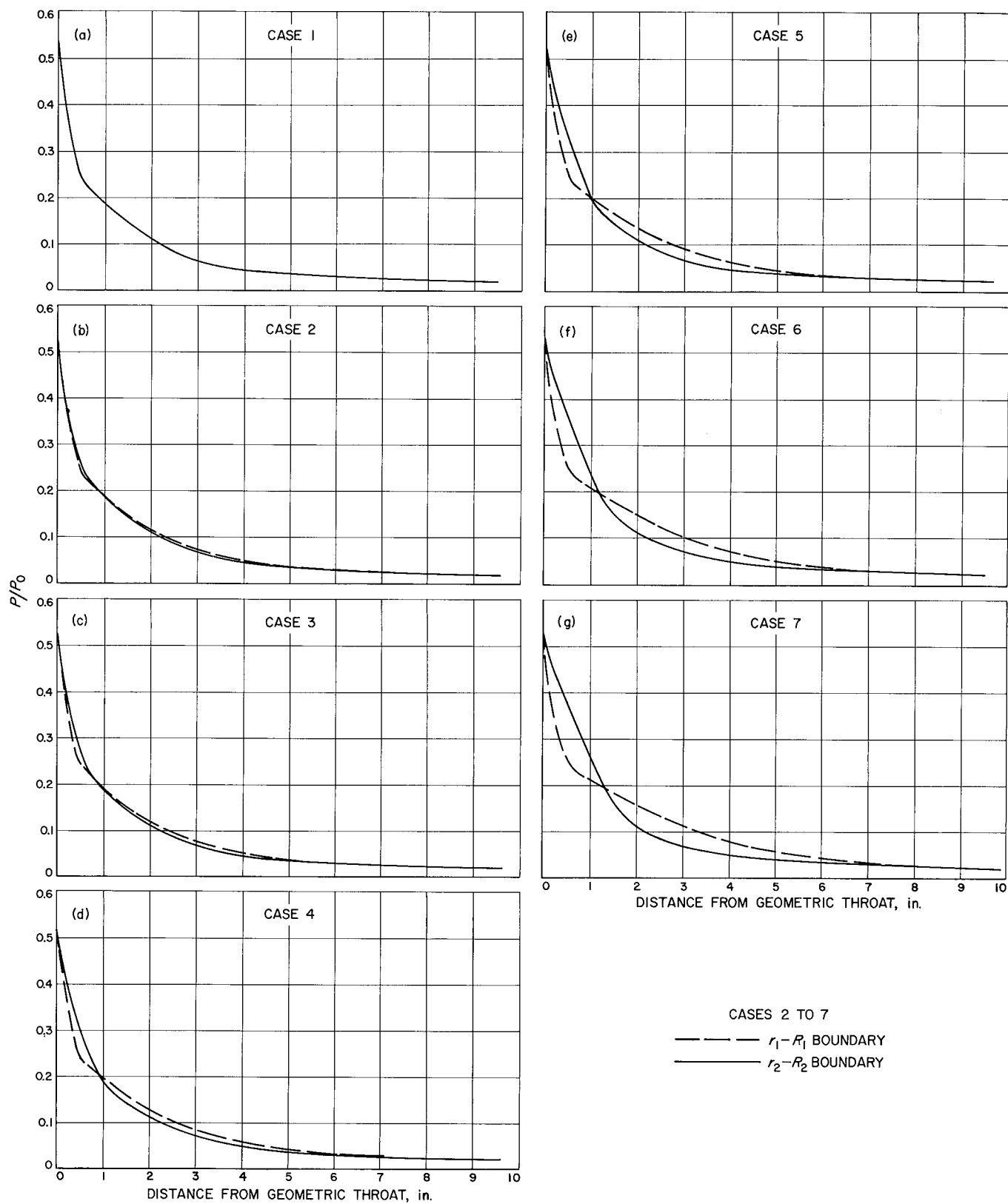


Fig. 4. Calculated wall static pressure ratio versus nozzle axial distance (a) Case 1, (b) Case 2, (c) Case 3, (d) Case 4, (e) Case 5, (f) Case 6, and (g) Case 7

agreement was considered adequate for qualitative-trend-type calculations, so that the second starting Mach line procedure was used for all calculations.

4. Results

The nozzle boundary surface pressure ratios calculated for the seven two-dimensional nozzle cases are shown in Fig. 4. The increasing divergence of the pressure profiles for the two boundaries with increasing nozzle asymmetry is evident. The measured cross-sectional static pressure ratio data for the asymmetric test nozzle is plotted in the same manner as the calculated results in Fig. 5. A comparison of Figs. 4g and 5 shows the expected differences between the pressure expansion profiles for the two- and three-dimensional nozzles. The greater area expansion with increased axial distance of the three-dimensional nozzle is accompanied by a much more rapid pressure expansion. The resulting differences in the pressure differentials and location of the crossover point of the two pressure profiles for the two different nozzles would be expected to result in somewhat different side-force characteristics also.

The net side force normal to the nozzle axis was calculated for each of the two-dimensional nozzle cases by numerically integrating the calculated pressure distributions over the two boundaries of each nozzle. A nozzle width of unity was used to simplify the calculations. The resulting side-force axial profiles, presented as the net side force/axial thrust ratio summed over the nozzle ex-

pansion ratio, for the seven nozzle cases are shown in Fig. 6. The nozzle axial thrust was calculated, assuming one-dimensional isentropic flow and neglecting the nozzle exit–ambient pressure differential term.

5. Discussion and Conclusions

An oscillatory type of side-force axial profile was obtained for each of the calculated cases. The summed net side-force ratio reached a maximum value; reversed itself at the point where the two pressure ratio profiles crossed; crossed the abscissa, reversing direction; and finally leveled off at the overall value for the nozzle. The peak and overall side-force values (amplitudes of the side-force axial profile) decreased with decreasing nozzle asymmetry. The crossover at the abscissa occurred at nozzle ϵ values of 2 to $2\frac{1}{2}$ (axial distances of 2 to 4 in. from the geometric throat). The overall side-force values decreased from a maximum of $2\frac{1}{4}\%$ of the axial thrust down to a value of approximately $3/4\%$ for the minimum asymmetry case (Case 2).

The calculated results obviously cannot be directly related to three-dimensional actual nozzle conditions, the pressure profile oscillation occurring at much smaller nozzle expansion ratios in the calculated cases as a result of the previously mentioned differing nozzle expansion characteristics, but certain trends and predictions can be arrived at.

As was previously pointed out, in the experimental test (Fig. 2) the side-force axial profile leveled off at a positive value, never crossing the abscissa. A nozzle with a smoother, less abrupt throat region contour (similar to the two-dimensional nozzle contours used), but with the same degree of asymmetry, would be expected to produce a smoother, less abrupt pressure expansion along the biased wall portion of the nozzle throat region, as illustrated by the dotted line in Fig. 5. This would result in a reduced maximum net side-force value and an overall value that approaches zero and probably crosses the abscissa, reversing the direction of force. Based on this interpretation of the experimental results, the effect of reduced magnitude of asymmetry on the net side-force characteristics for three-dimensional nozzles is predicted to be qualitatively the same as for the two-dimensional calculated results, a reduction in the amplitudes of an oscillatory type of net side-force axial profile. The overall nozzle side force should be less than 1% of the nozzle axial thrust ($\frac{1}{2}$ deg of misalignment in the thrust vector) for the magnitudes of nozzle throat asymmetry experienced in actual rocket nozzles.

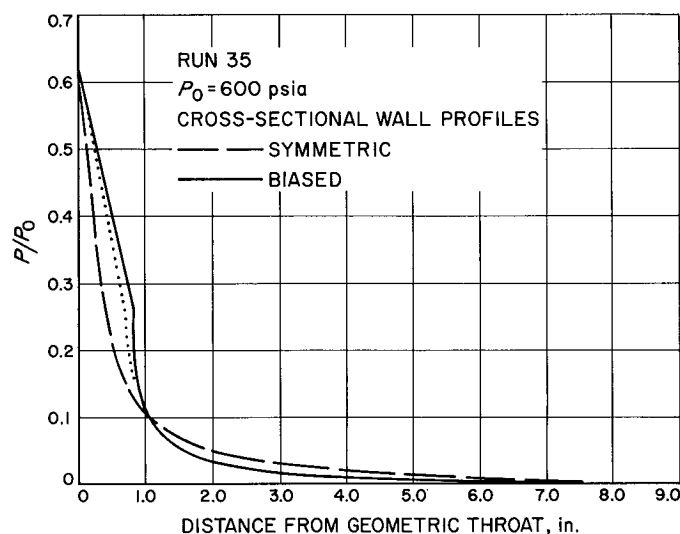


Fig. 5. Measured wall static pressure ratio versus nozzle axial distance, asymmetric nozzle

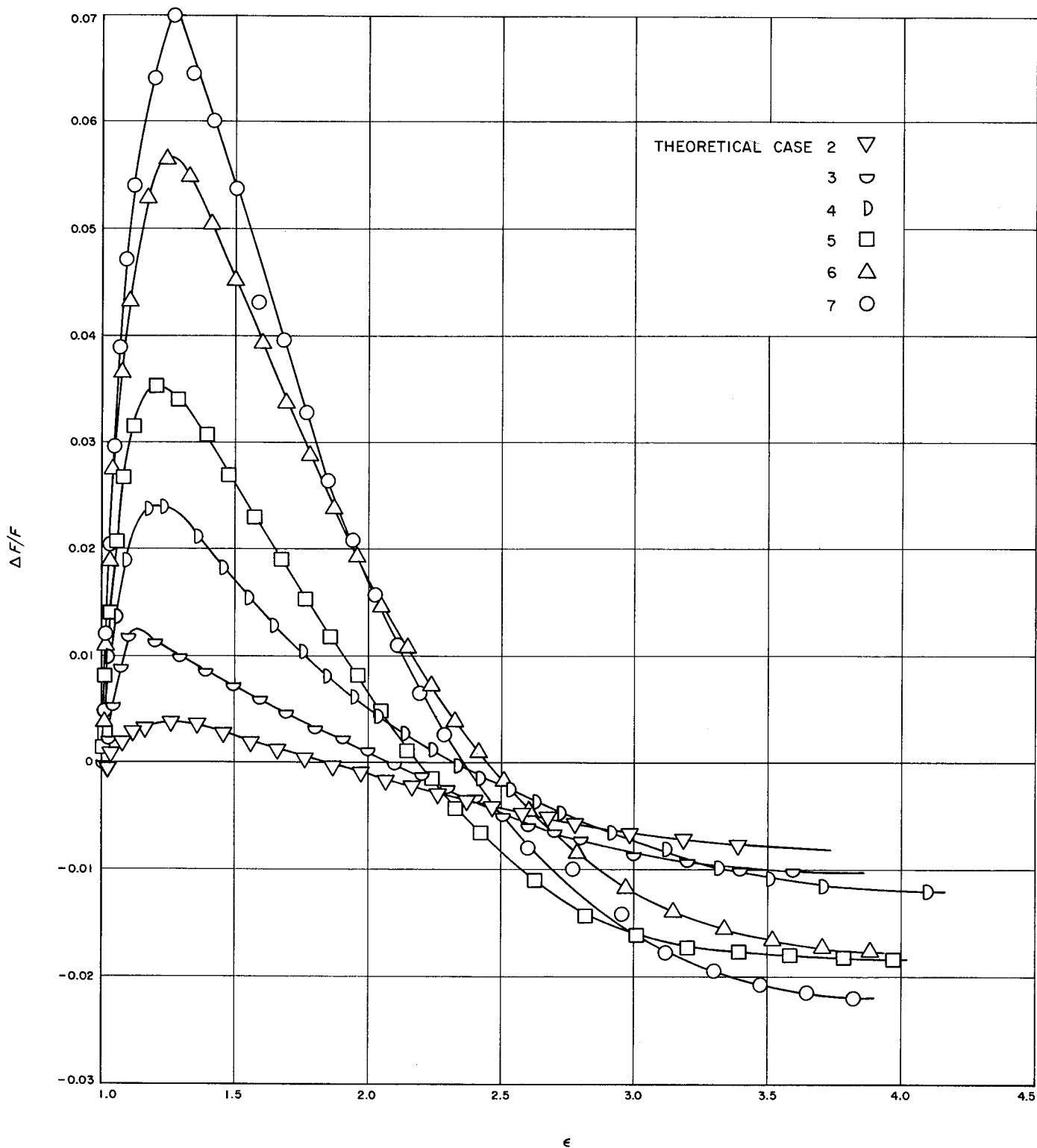


Fig. 6. Summed net side force/axial thrust ratio versus nozzle expansion ratio, two-dimensional asymmetric nozzles

The theoretical analysis contained in a recent paper on this subject (Ref. 2) also predicts an oscillatory type of side-force axial profile as a result of asymmetric flow in the throat region of a rocket nozzle. The paper also reports the results of some rocket static-firing experiments that are in general agreement with the preceding conclusions. Small rockets with nozzles with known asymmetries were static-fired in a six-component test stand. A reported curve of lateral force/axial thrust versus nozzle length exhibited the oscillatory form, with an amplitude of 0.5%. A few similar small rocket test firings are currently being planned to attempt to verify the experimental and theoretical results and the conclusions presented here.

References

1. Ehlers, F. E., "An IBM 7090 Program for Computing Two-Dimensional and Axially-Symmetric Flow of an Ideal Gas," *Mathematical Note No. 727*, Mathematics Research Laboratory, Scientific Research Laboratory D 1-82-0204, Boeing Co., Seattle, Wash., November 1962.
2. Darwell, H. M., and Trubridge, G. F. P., "The Design of Rocket Nozzles to Reduce Gas Misalignment," *Preprint of ICRPG/AIAA 2nd Solid Propulsion Conference*, pp. 198 to 205, Anaheim, California, June 1967.

C. Prepolymer Functionality Determination Using a Model Polymerization System, H. E. Marsh and J. J. Hutchison

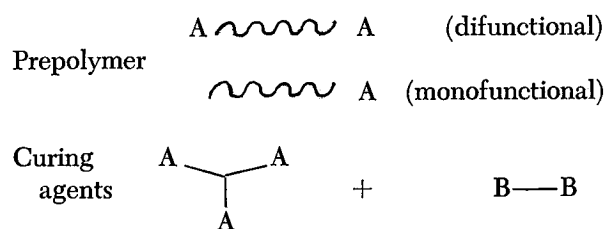
One of the chief aims of an investigation of a series of polymerization reactions reported earlier (SPS 37-42, Vol. IV, p. 106; SPS 37-43, Vol. IV, p. 163; SPS 37-45, Vol. IV, p. 77; SPS 37-47, Vol. III, p. 69) was the development of a model polymerization system suitable for the determination of functionality in prepolymers. The functionality of prepolymers is defined as the average number of reactive groups (reactive in the sense of being able to unite with complementary reactive groups on other constituents in order to effect the process of polymerization) per molecule. The method of functionality determination in current use is a combination of reactive group assay and molecular weight measurement. It is indirect, and its accuracy is unreliable. None of the reactions with the desirable feature of having no elimination products was found to also have the more necessary quality—no significant side reactions. For this reason, the reaction finally chosen for the model system was esterification. As was reported previously (SPS 37-47, Vol. III, p. 69), esterification appears to fit the basic requirements of high yield and lack of significant side reactions. Esterification has also a distinct additional advantage: it can be employed

directly with all prepolymers of current interest because it involves both of the two reactive groups in use today: carboxylic acids and aliphatic hydroxyls. This polymer-linking reaction does have two disadvantages; however, data obtained during this reporting period indicate that these problems are under control as a result of refinements in technique.

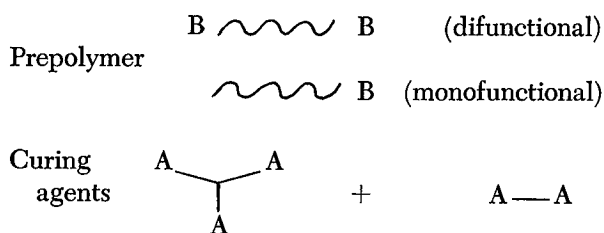
The two problems associated with applying esterification to functionality determination, in the manner being considered here, are related by virtue of their effects on high yield. First, esterification is slow; considerable time is required in the case of a given sample before it is safe to assume that no further reaction will take place. Second, esterification is an equilibrium reaction; the elimination product, water, must be removed completely to shift the equilibrium reaction toward completion. The use of temperatures in the range of 140 to 170°C accelerates the reaction and promotes water removal as well. Even at that, complete reaction takes several weeks. Catalysts are now being studied as another means of reducing the necessary time. In earlier experiments, a small degree of sample discoloration was observed, indicating some degradation. This was assumed to be caused by oxygen in the air and the high temperature. The use of a nitrogen atmosphere eliminated this phenomenon, verifying the hypothesis. Continual replacement of the nitrogen atmosphere with dry nitrogen serves to carry away water produced by the reactions, thus helping with the equilibrium shift problem. If the water removal appears to be a continuing problem, vacuum may be applied during the later stages of reaction, after the more volatile low-molecular components will have been built into the polymer. Additional comments on equilibrium and water, in the following discussion, will show an additional advantage of the approach under investigation.

The theory behind the subject method of functionality analysis is based on calculation of probabilities of the formation of infinite networks in condensation polymerization. The reasoning of Flory (Ref. 1) is followed. Two forms of model polymer systems are discussed here as being most useful for functionality determination.

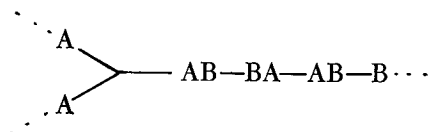
Type I



Type II



A and B are reactive groups which unite to form linkages. Thus, the polymer network is built, as the reaction of A with B proceeds.



The component of special concern illustrated in these diagrams is, of course, the prepolymer. Ideally, prepolymers are difunctional. Practically, this is not often the case. Prepolymers in use, or under study, are synthesized from small monomer units to form chains ranging in average molecular weight from 500 to 6000. No matter what monomer unit or synthetic method is used, products are mixtures containing distributions of molecular size and distributions of di-, mono- and zero-functionality. The presence of zero- or nonfunctional molecules is important in respect to the properties of elastomers made from the prepolymers, but this factor does not enter into the probabilities relating to network formation. Some fraction of the molecules in some prepolymers have functionalities higher than 2; however, the present treatment does not handle this possibility. It will be taken up at a later time. For the purposes of this discussion, functionality will be defined as follows:

f = the average number of reactive groups per molecule, excluding all zero-functional molecules.

For the systems under discussion, in which the prepolymers are mixtures of only two functionalities, di- and monofunctional (again, nonfunctional molecules may be present, but do not enter into the network),

F = the mole fraction of difunctional molecules

$$F = 2 \left(1 - \frac{1}{f} \right), \text{ and } f = \frac{2}{2-F} \quad (1)$$

Theoretically, it is possible to convert a balanced mixture of di- and trifunctional components to one very large

molecule having a structure resembling three-dimensional chicken wire, with the spacing between branch sites, or cross-links, regulated by the concentration of the trifunctional components. Clearly, monofunctional molecules, if present, will terminate some of the growing chains and produce some low-molecular-weight molecules, without necessarily preventing the formation of an infinite network. When this occurs, part of the product is gel, and part is sol. Similar results will be caused by other characteristics of the system—lack of stoichiometric balance of A and B reactive groups and lack of complete reaction. Flory (Ref. 1) introduced a term, branching probability, or branching coefficient α and showed its dependency on the latter two characteristics, when no monofunctional molecules are present, to be

$$\alpha = \frac{P_A^2 \rho_T}{R - P_A^2 (1 - \rho_T)} \quad (2)$$

or

$$\alpha = \frac{P_B^2 \rho_T}{r - P_B^2 (1 - \rho_T)} \quad (3)$$

where

α = the probability of starting with one of the reactive groups A of a randomly chosen branching unit (trifunctional in this case), proceeding via a chain of connected difunctional units (of alternating polarity, B—B, A—A, etc.) and reaching another branching unit

ρ_T = the fraction of A groups contributed by branching units

R = ratio of equivalents of B to equivalents of A

$r = 1/R$

P_A, P_B = fraction of A and B groups, respectively, that have reacted.

Since R = (concentration of B)/(concentration of A), then also $R = P_A/P_B$. The same, of course, is true of r , and it is further observed that the symmetry shown in Eqs. (2) and (3) holds for all the following (Eqs. 4 and 5). Generally, it is convenient to use the equation in which the stoichiometric ratio, R or r , is greater than one.

In an earlier publication (Ref. 2) we showed that monofunctionality could be accounted for. The equivalent

expression for the Type I model derived from probabilities is

$$\alpha = \frac{P_A^2 \rho_T}{R - P_A^2 F(1 - \rho_T)} \quad (4)$$

The corresponding expression for the Type II model derived by us recently is

$$\alpha = \frac{P_A^2 F \rho_T}{R - P_A^2 F(1 - \rho_T)} \quad (5)$$

These new relations are in agreement with more general equations developed by Stockmayer (Ref. 3) and Kahn (Ref. 4). It should be pointed out that the functionalities in Stockmayer and Kahn's equations are weighted averages; those used in this report, as in Flory (Refs. 1 and 5), are number averages.

Flory (Ref. 5) reasoned that, in the case of a trifunctional branching unit, n chains can be expected to lead, by tracing them individually, to $2n\alpha$ chains. If $2n\alpha$ is greater than one, n chains lead to more than n chains, and the branching can be expected to go on indefinitely; an infinite network or gel exists. Conversely, if $2n\alpha$ is less than one, n chains lead to fewer than n chains, and eventually terminate. Thus, a value of $\frac{1}{2}$ for α is critical where the branching unit is trifunctional. It is the condition for incipient gelation; $\alpha_c = \frac{1}{2}$. Figures 7 and 8 illustrate the full range of incipient-gelling compositions (assuming complete reaction, $P = 1$) for Types I and II systems. These are plots of Eqs. (4) and (5). Notice the greater sensitivity of Type II systems to functionality of the prepolymer.

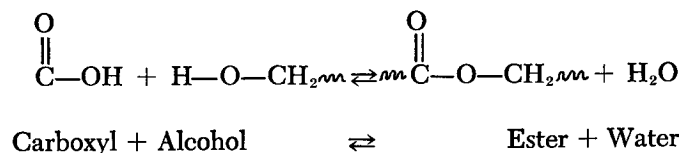
The essential plan of this method of functionality determination is to find which of an array of polymer formulations having slightly varying composition marks the composition of incipient gelation at complete reaction. Prepolymer functionality is calculated from the composition (R and ρ_T) of this critical formulation using Eq. (4) for Type I or Eq. (5) for Type II and Eq. (1), giving α the value of one-half and P_A (or P_B) the value of one. Preliminary data with purified (SPS 37-43, Vol. IV, p. 163) Dimer acid (from Emery Industries) as a model prepolymer are in reasonable agreement with functionality obtained from molecular weight and carboxyl measurements. Work is under way on other prepolymers.

A method of prepolymer functionality determination somewhat similar to the one described here was reported by Strecker and French (Ref. 6). Incipient gelation was used as the analysis end-point; however, with their approach, it is necessary to arrest the reaction and analyze for

unreacted groups by other means. The methods are different in still another way. Strecker and French used only stoichiometric ($R = 1$) mixtures; we expect to gain additional useful information by varying the stoichiometry.

It is planned to investigate some other aspects of this analysis scheme. These can be illustrated by referring to Figs. 7 and 8, which are, for Types I and II, respectively, families of constant functionality curves at complete reaction. The first thing to notice is that in each case the relation between R and ρ_T is linear. Thus, if complete reaction is attained, identification of two incipient-gelling compositions (two sets of R and ρ_T) provides data for the determination of another characteristic of the system besides prepolymer functionality, such as the equivalent weight of the prepolymer.

The second area to be investigated has to do with the attainment of complete reaction and the effect of equilibrium on that attainment. As was mentioned before, the elimination product, water, must be completely removed to shift the reaction to completion. Other factors that also have a practical bearing on this matter, such as reaction kinetics and the mobility of reactive groups in an ever increasingly viscous liquid will be ignored for the time being. Symbolically, the esterification reaction is as follows:



The equilibrium relation for this reaction in bulk is

$$K = \frac{[\text{ester}][\text{water}]}{[\text{carboxyl}][\text{alcohol}]} \quad (6)$$

in which the brackets [] signify mole fractions. For the simple system, acetic acid, ethanol, ethyl acetate and water, the equilibrium constant K is 4.0, and it does not change greatly with temperature (Ref. 7). Although the equilibrium constants of the various polymer systems in this study are not likely to be the same as for acetic acid and ethanol, a value of 4.0 is considered to be close enough for examination of the effects of dissolved water and equilibrium on this analysis.

The very low concentration of water needed to approach complete esterification is illustrated in Fig. 9; these curves were calculated from Eq. (6), using $K = 4.0$. The figure

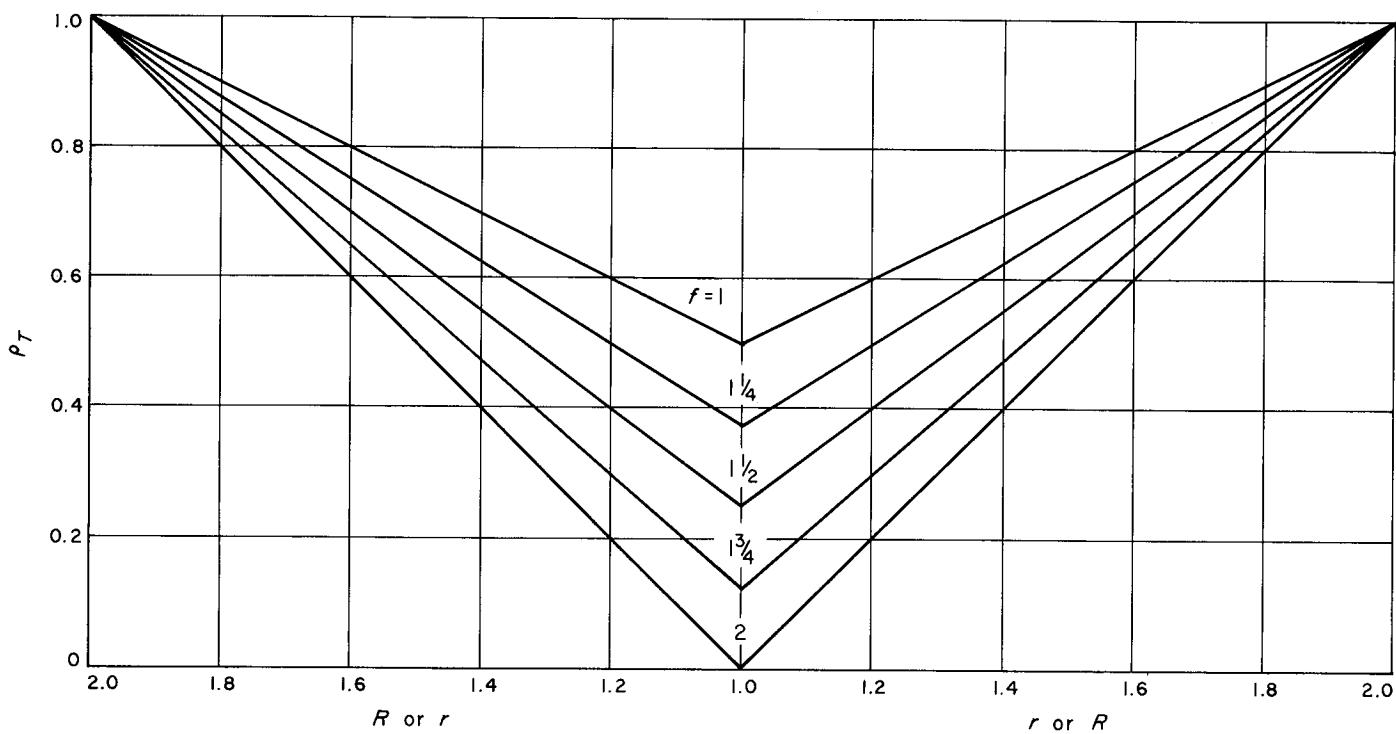


Fig. 7. Incipient gelation lines, Type I polymerization system

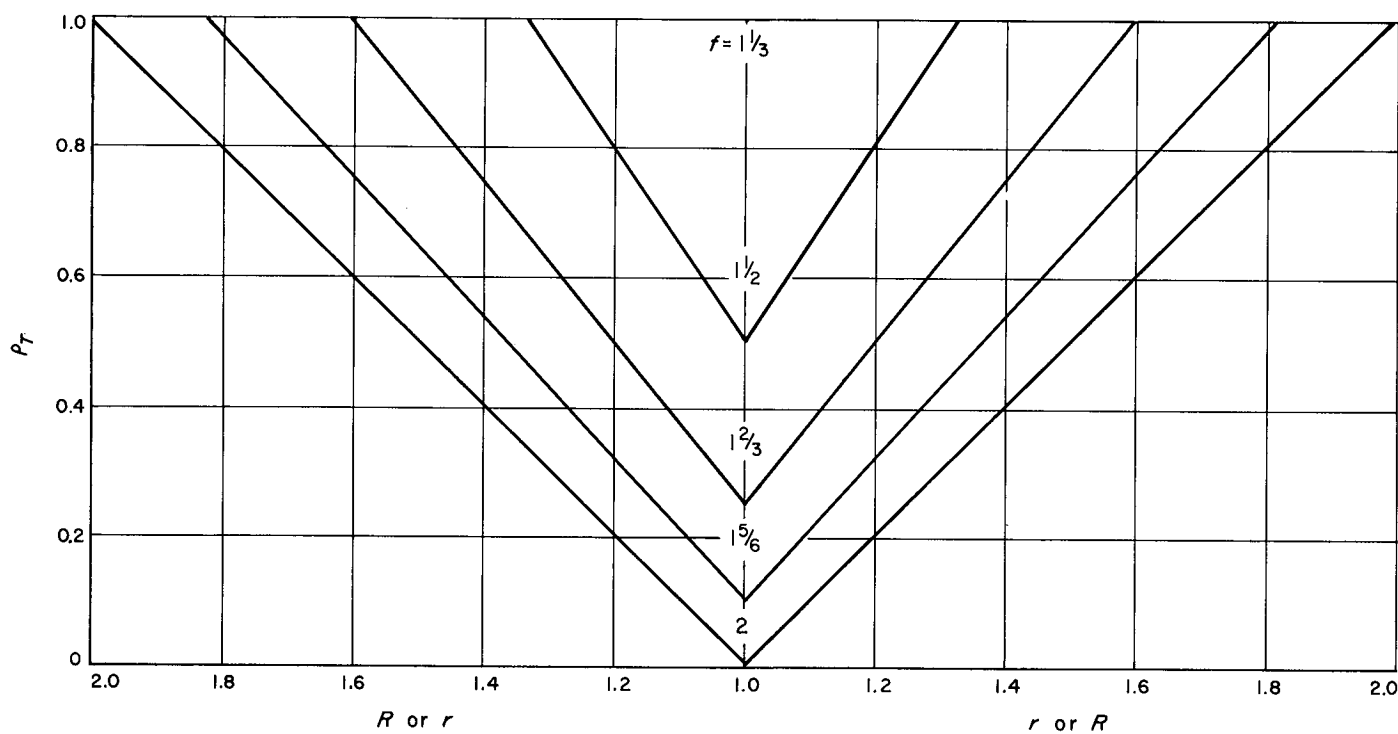


Fig. 8. Incipient gelation lines, Type II polymerization system

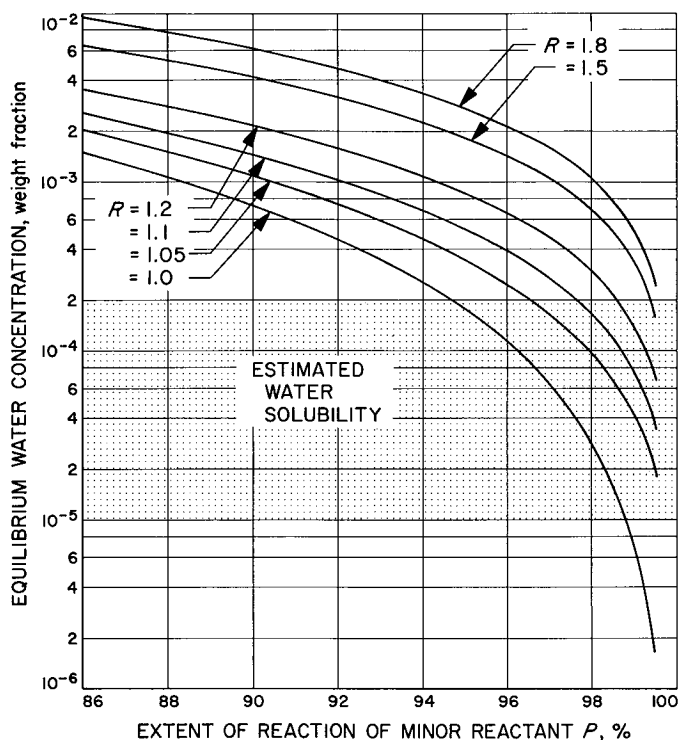


Fig. 9. Water tolerance of esterification reaction

shows that a large advantage is to be gained by using nonstoichiometric mixtures of reactive groups ($R \neq 1$). If one of the reactants is in excess, the other, or minor, reactant tends more nearly to be used up; e.g., shifting R from 1.0 to 1.05 permits an almost tenfold increase in water tolerance to achieve a conversion of $P = 0.995$. Currently, experiments are being run with nonstoichiometric mixtures.

It was mentioned above that the linear relation between R and ρ could be exploited to obtain prepolymer reactive group assay as well as functionality, if complete reaction could be assumed. However, if complete reaction is not obtained, then the system can be approached from another angle in order to effectively measure the true values of prepolymer functionality. If we can assume, instead of

complete reaction, that equilibrium is reached and that equal water concentration is reached in a given set of samples exposed to the same environment, the data from two incipient-gelling compositions can be treated simultaneously in equilibrium (Eq. 6) and probability (Eq. 4 or 5) relations to solve for functionality and extents of reaction. This will be investigated further.

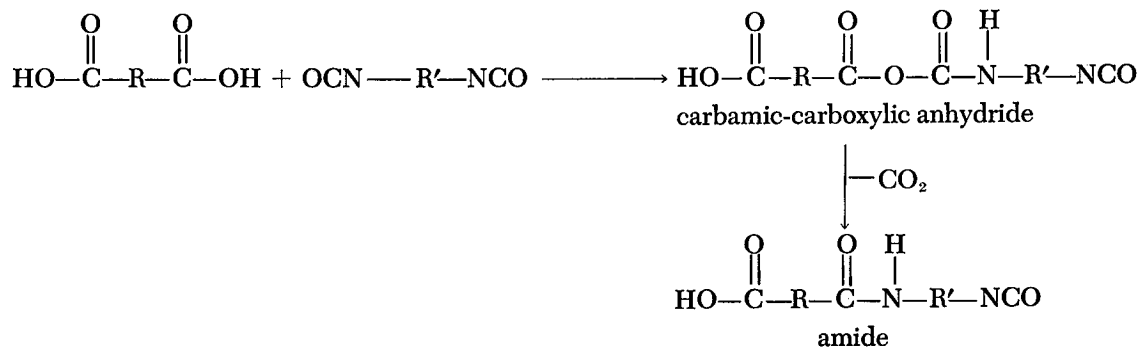
References

1. Flory, P. J., *J. Am. Chem. Soc.*, Vol. 63, p. 3083, 1941.
2. Marsh, H. E., Jr., *Ind. & Eng. Chem.*, Vol. 52, p. 768, 1960.
3. Stockmayer, W. H., *J. Pol. Sci.*, Vol. IX, No. 1, p. 69, 1952.
4. Kahn, A., *J. Pol. Sci.*, Vol. XLIX, p. 283, 1961.
5. Flory, P. J., *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, New York, 1953.
6. Strecker, A. H., and French, D. M., *Polymer Preprints*, Vol. 7 (2) ACS Div. of Polymer Chemistry, p. 952, September 1966.
7. Glasstone, S., *Textbook of Physical Chemistry*, D. Van Nostrand, New York, 1946.

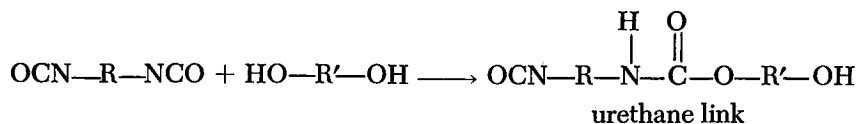
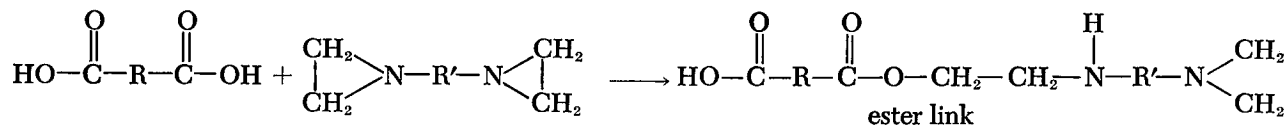
D. Foams Produced From Carboxyl-Terminated Hydrocarbons, S. Anderson, J. J. Hutchison, and H. E. Marsh, Jr.

Elastomeric foams which can withstand elevated temperatures without significant decomposition or loss of mechanical properties have many potential uses. A previous study (SPS 37-36, Vol. IV, p. 154) investigated the feasibility of foams for use as liners in solid propellant motors which are to be heat sterilized. In that study the foams were based on carboxyl-terminated hydrocarbon prepolymers cured with aziridines and used commercial blowing agents.

In the present work, a different curing and gas-producing reaction was investigated. The reaction of carboxyl-terminated prepolymer with a diisocyanate will not only provide carbon dioxide for foam formation but will yield amide chain-extending linkages.



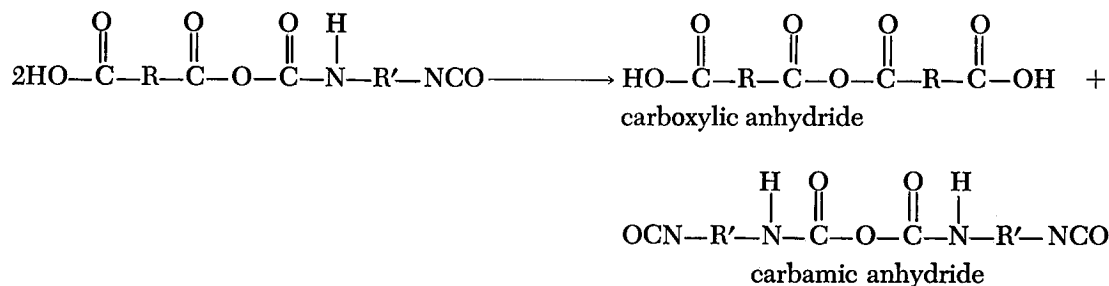
The inclusion of a trifunctional carboxyl-terminated hydrocarbon will produce a crosslinked network which will give the foam its necessary rubbery properties. The amount of gas, and thus the foam density, could be regulated by adding reagents which would react with either the carboxyl or isocyanate groups to give chain-extending linkages but no gaseous elimination products. Examples are multifunctional aziridines which produce ester linkages and multifunctional alcohols which yield urethanes.



Optimization of a number of critical variables is necessary for producing useful foams. If the prepolymer cures before sufficient gas has evolved, the viscosity will be too high for a foam to rise. If gelation takes place too slowly the gas will escape, and the foam will collapse. The variables and ingredients considered here were:

- (1) Choice of carboxyl-terminated prepolymer.
 - (a) Unsaturated Telagen (General Tire and Rubber Corp.).
 - (b) Saturated Telagen (General Tire and Rubber Corp.).
 - (c) Dimer acid (Emery Industries).
 - (d) Dimer acid polyester (SPS 37-47, Vol. III, pp. 73-74).
- (2) Amount of trifunctional carboxyl crosslinker—Trimer acid (Emery Industries).
- (3) Amount and choice of diisocyanate.
 - (a) Hexamethylene diisocyanate.
 - (b) Toluene diisocyanate.
- (4) Amount, if any, of a difunctional aziridine, HX 740 from 3M Co.
- (5) Amount, if any, of a saturated hydroxyl-terminated prepolymer, Telagen (General Tire and Rubber Corp.).
- (6) Amount of reaction catalyst, triethylenediamine.
- (7) Amount of cell-control agent, silicone oil DC 200 from Dow Corning Co.
- (8) Foaming temperature and time.
- (9) Curing temperature and time.

The use of toluene diisocyanate was found early to produce very poor, sticky, weak, and collapsed foams. Analysis by infrared spectrophotometry showed the presence of significant amounts of carboxylic acid anhydride linkages. These are the result of an unwanted rearrangement of the mixed carbamic-carboxylic anhydride intermediate.



Such a rearrangement prevents proper curing and carbon dioxide evolution. It was minimized by using an aliphatic diisocyanate, hexamethylene diisocyanate, rather than the aromatic toluene diisocyanate.

The addition of triethylenediamine catalyst in the range 0.5 to 1.1 wt % was found necessary to give good foaming rates. However, an excess of catalyst was harmful to foam structure during curing. The poor solubility of the solid catalyst at room temperature, even after grinding to a powder, was also a problem because of difficulty in obtaining a homogeneous mixture.

The silicone oil emulsifier improved the formation of cells in the foam. It was soluble in the reaction mixture, but when used in amounts greater than about 1.0 wt % it remained as an oily film on the cured foam.

Unsaturated carboxyl-terminated Telagen was clearly inferior to the saturated as shown by the collapse of its foams at a curing temperature of 100°C and weakness even at room temperature. Dimer acid was also eliminated as a prepolymer by its low molecular weight. Its resulting high carboxyl content per gram (3.4 meq as opposed to 0.8 for Telagen) produced too much carbon dioxide for useful foam. Dimer acid polyester with a molecular weight similar to that of Telagen gave results as good or even slightly better than those of the saturated Telagens. This polyester has the advantages of lower viscosity, which facilitates mixing, and more nearly approaching the ideal of difunctionality. The available Telagens have approximately one quarter of their chains with one nonfunctional end. Therefore, the final cured polymer network is weakened.

The critical factor in determining the foam strength was the inclusion of Trimer acid to produce necessary crosslinks, forming a three-dimensional-polymer network. Very high proportions of Trimer acid (ratio of equivalents of carboxyl from Trimer acid to total carboxyl equivalents equal to 0.8 to 0.9) were required to give even moderately tough flexible foams, no matter what the prepolymer. This is a good indication of incomplete reaction of carboxyl groups, because such a large amount of crosslinker will normally produce very rigid, brittle material.

Trimer acid has a carboxyl content per gram as high as that of dimer acid; so reagents to reduce carbon dioxide evolution were necessary. Hydroxyl-terminated Telagens were not very effective. Apparently the difference in hydroxyl-isocyanate and carboxyl-isocyanate reaction

rates hinders their use in combination. Much better results were achieved with the diaziridine HX 740. Foam density and toughness increased as the ratio of equivalents of aziridine to equivalents of carboxyl was increased through a range of 0.2 to 0.8. The ratio of isocyanate equivalents to carboxyl correspondingly decreased from 0.8 to 0.2. Using an excess of isocyanate and aziridine to carboxyl gave no improvement.

Reactants were generally mixed at room temperature and foamed without external heating. However, there was sufficient heat evolution from the aziridine-carboxyl reaction to bring the center of a 1.5-in.-diam foam to about 45°C for 10 min. The foams were allowed to set for around 20 h and were then cured at 85°C for 1 to 3 days. Increasing the cure temperature or time did not improve properties.

The result of this study was a series of foams with typical densities between 0.13 and 0.28 g/cc. The best of them were sticky and flexible, with fairly uniform cell structure. As foam density decreased, the cells tended to be more jagged and uneven. All the foams either dissolved or softened in hexane, although high trimer acid and HX 740 content definitely improved their resistance.

Their most important property, high-temperature stability, was disappointing. A period of 90 h at 120°C resulted only in darkening, but at 135°C only short exposures could be tolerated before loss of strength became obvious.

We conclude that, although the carboxyl-isocyanate reaction can successfully produce foams, the system contains some inherent problems which must be solved before heat-resistant foams can be obtained.

Two of these are presumed to be:

- (1) Unreacted isocyanate groups which exist after the foam sets. This is the same as in urethane foams. However, in urethanes the gas-producing reactions are essentially complete at this point, and curing only strengthens the polymer network. In the carboxyl-isocyanate system, curing also produces more carbon dioxide which may damage cell walls.
- (2) Even after curing, the network forming reactions are not complete, as shown by stickiness, hexane solubility, with crosslinker ratios as high as 0.9. This may be a result of poor reaction rates or side reactions, such as the rearrangement of the mixed carboxylic carbamic acid anhydride intermediate as was shown to occur with toluene diisocyanate.

E. Transition From Deflagration to Detonation in Granular Solid Propellant Beds, O. K. Heiney

1. Introduction

As part of an experimental program to verify a developed analytic interior ballistic formalism described in SPS 37-43 and 37-44, Vol. IV, a series of instrumented ballistic firings were conducted. In the main, the correlation between theory and experiment was excellent; however, during the course of the firings a very interesting phenomenon was uncovered. It was noticed that as propellant charges were increased relative to loading volume, a transition from normal deflagration to shock-driven deflagration to total detonation was encountered. This effect appears to be correlatable in terms of a function of the

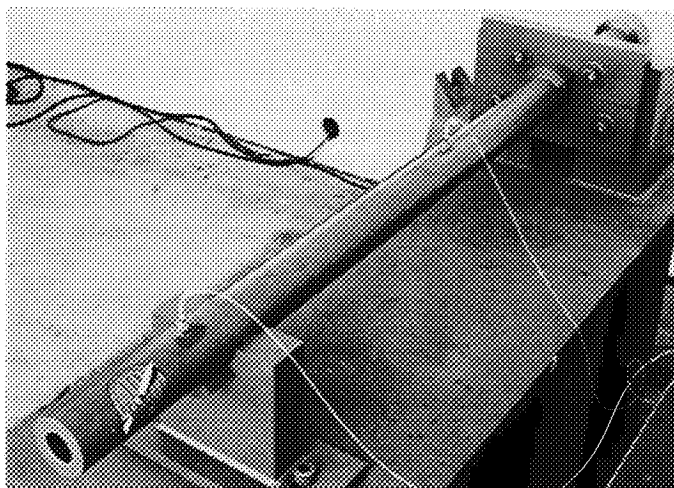


Fig. 10. Firing device (37 mm)

propellant burning surface per unit free initial packing volume.

2. Discussion

a. System configuration. The basic function of a gun ballistic theory is to predict the pressure-time history of any arbitrary system, given only the dimensions and physical characteristics of the system, the projectile, and the propellant. The validity of the predictions is easily assessed by then measuring the pressure history of a given device and correlating the results with the theory. The device which was used for this purpose is illustrated in Fig. 10.

Pressures were measured in the chamber and at two additional points down the barrel by means of high-pressure Kistler piezometric pressure transducers feeding Kistler charge amplifiers and recorded on persistent phosphor-type Tektronix oscilloscopes. Muzzle velocities were measured by means of break screens connected to Hewlett Packard digital clocks.

The launch tube was of smooth bore configuration and for maximum flexibility was constructed with a uniform bore diameter rather than with an expanded chamber. This design allows an infinitely variable chamber volume. Ignition and propellant loading techniques are illustrated in Fig. 11.

The initial chamber volume is determined by the location of the piston base, while the loading volume is a function of the diameter and length of the phenolic sleeve into which the propellant is initially packed. The igniter used consists of a firing nut containing an Atlas electric

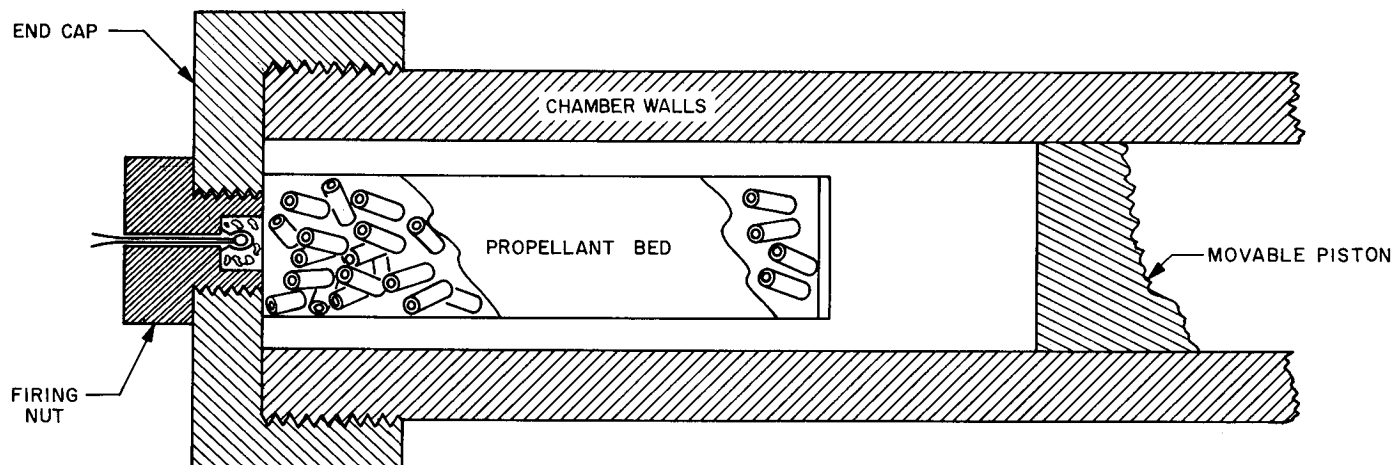


Fig. 11. Breech and ignition arrangement

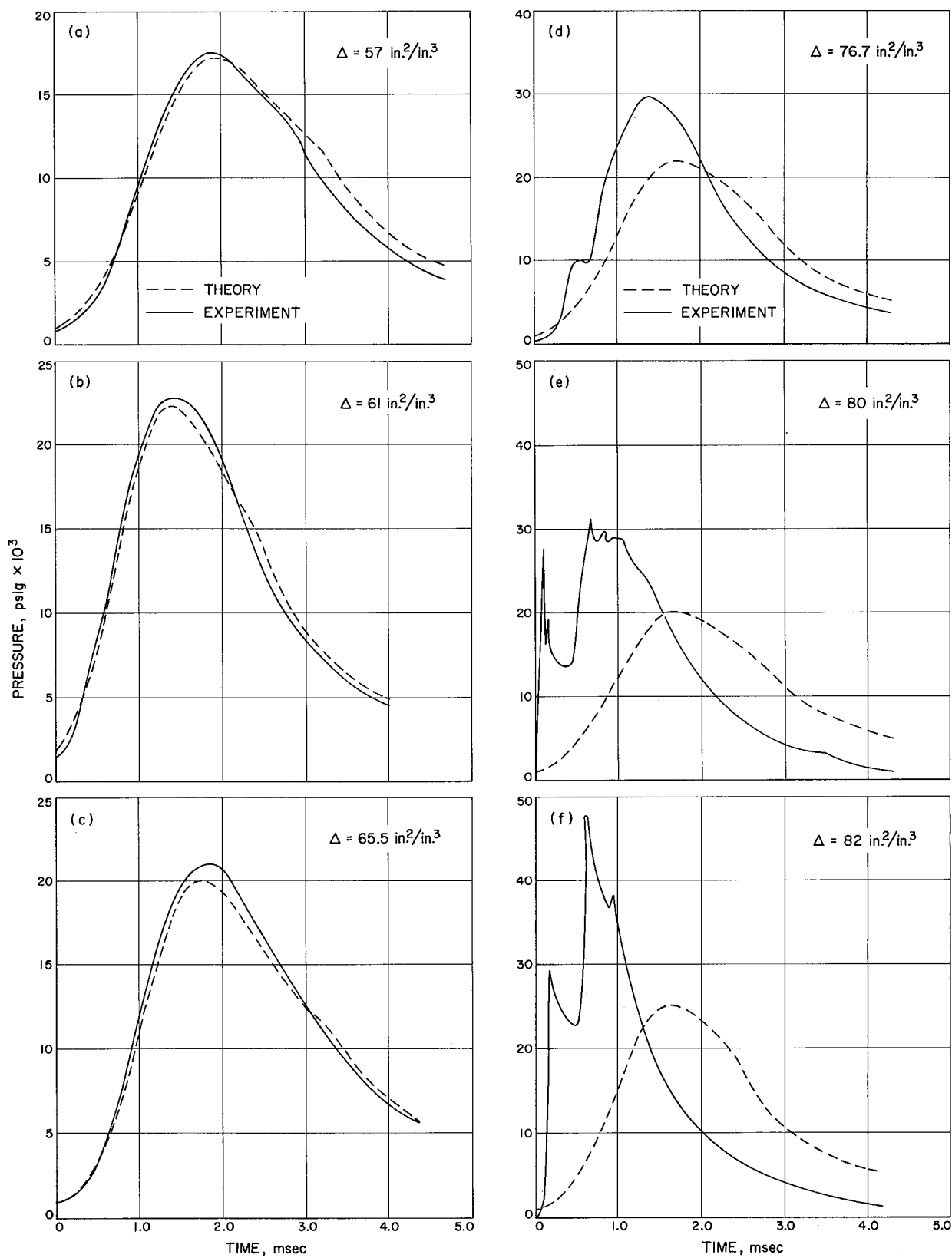


Fig. 12. Pressure-time plot for ballistic device

match surrounded by 300 mg of black powder. This type of igniter gives a short-duration high-intensity flame with little brisance. When the tubular propellant grains are not too tightly packed, the effect of the phenolic loading sleeve may be neglected except insofar as the volume it displaces is concerned.

b. Deflagration characteristics. Firings demonstrating proper deflagration and typical correlation with the above-referenced theory are illustrated in Figs. 12(a), (b), and (c). A qualitative delineation of the phenomena occurring in Fig. 12(a), for example, would be as follows:

- (1) $t = 0$ to 1.0 ms. Very slow increase in chamber volume due to almost negligible projectile velocity, hence very rapid pressure increase due to energy release by propellant in almost constant chamber volume.
- (2) $t = 1.0$ to 1.9 ms. Projectile velocity increasing and thus exposed chamber volume increasing more rapidly. Excess energy input decreasing as function of incremental volume to be pressurized.
- (3) Peak pressure ($t = 1.9$) to propellant burnout ($t = 2.9$ experimentally, 3.2 analytically). Plenum volume increasing more rapidly than energy input. Sharp break in curve slope due to propellant burnout.
- (4) Subsequent to burnout a very rapid pressure decrease occurs, due to the expansion of the gases, heat loss to tube, and further energy imparted to projectile.

In these three firings, the primary difference exhibited between the experimental data and the analysis is due to the following factor: the single perforate propellant (Fig. 13) is assumed to burn externally and internally in a radial manner until the total charge is consumed.

In reality, this does not occur. A certain fraction of the grains fragment during combustion; this increases the

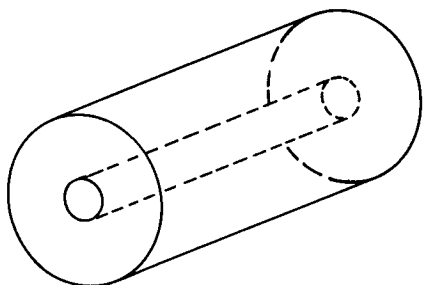


Fig. 13. Propellant grain

exposed burning surface and, in general, leads to slightly higher peak pressures and somewhat earlier web burnout than is analytically predicted. The difference typically is of only a few percent in peak pressure and is of such a nature that an attempt to express it analytically would probably be futile.

Figures 12(a), (b), and (c) then illustrate the quite good correlation between theory and experiment which is attainable when the density of loading Δ , more fully described later, is kept at less than 75 in.²/in.³. At loadings above this level a different mechanism of combustion becomes manifest and could probably be best described as a transition through shock-driven deflagration to total charge detonation.

The characteristics of the six illustrated firings are as delineated in Table 6.

Table 6. Firing characteristics

Figure No.	Charge weight, g	Propellant web, in.	Slug weight, g	Δ , in. ² /in. ³	Muzzle velocity, ft/s	
					Predicted	Actual
12 (a)	86.2	0.0164	560	57	2080	2040
12 (b)	90.0	0.0164	560	61	2220	2210
12 (c)	110.0	0.0190	508	65.5	2360	2300
12 (d)	97.5	0.0164	560	76.7	2250	2150
12 (e)	93.0	0.0164	550	80	2190	2220
12 (f)	105.0	0.0164	560	82	2350	2700

c. Transition to detonation. It was mentioned above that the phenolic tube had no effect on the interior ballistic solution. This is not true, however, if the propellant is too tightly packed into the tube. The propellant used was M-10, which is virtually 100% nitrocellulose with no nitro-glycerine loading and hence would be expected to be relatively insensitive with regard to detonation characteristics.

Figures 12(d), (e), and (f) graphically display this phenomenon of transition from weak shock-driven deflagration to strongly shock-driven deflagration. Figure 14 is an illustration of the remains of an end cap, as shown in Fig. 11, when loading was increased to the point where complete detonation occurred. No pressure record is available for this firing as the breech pressure transducer was in a condition roughly comparable to that of the end cap. An indirect method of pressure determination may be made by the calculation of the force necessary to shear the normalized 4130 end cap and indicates a peak pressure of at least 450,000 psig.

From the illustrated firings and many others conducted, it is possible to characterize the combustion of the propellant into various regimes as a function of the burning surface per unit free initial volume.

Symbolically this would be

$$\Delta = S_B/V_{IF}$$

From SPS 37-43, Vol. IV, p. 167, for single perforate grains

$$S_B = 2 C_W/\rho_P \mathcal{W}_0$$

then

$$V_{IF} = L A - C_W/\rho_P$$

where

A = phenolic tube area

C_W = initial charge weight

L = phenolic tube length

S_B = propellant burning surface

V_{IF} = initial free chamber volume

\mathcal{W}_0 = propellant web

Δ = burning surface per unit free volume

ρ_P = propellant density in lb/in.³

It was mentioned above that for proper deflagration, in this system, Δ must be less than 75 in.²/in.³. In Fig. 12(d) $\Delta = 76.7$ in.²/in.³, and it is seen that a slight pressure pulse occurs, then damps out, but drives the peak pressure to a value approximately 30% higher than what would have been encountered during proper deflagration.

In Fig. 12(e), for which $\Delta = 80$ in.²/in.³, the initial pressure spike is rapid and narrow, quickly decays, and is followed by another broader pressure pulse to approximately the same value. Figure 12(f) illustrates the pressure-time profile of a loading with a $\Delta = 82$ in.²/in.³

The initial pressure wave is virtually identical to that exhibited by the previous loading; however, the second spike is much higher and gives a peak pressure almost 100% higher than would be expected if linear regression were the only mechanism at work.

The loading represented by the wreckage shown in Fig. 14 was $\Delta = 86.5$ in.²/in.³. Complete detonation of the propellant had taken place resulting in a pressure probably well above the 450,000 psig mentioned above.

From the above results for the system under consideration, the following regimes can be defined:

$$\Delta < 75 = \text{proper deflagration}$$

$$75 < \Delta < 86 = \text{shock-driven deflagration}$$

$$\Delta > 86 = \text{detonation}$$

These particular Δ values are doubtless a function of propellant composition and ignition technique. They do graphically demonstrate, however, that the ballistic designer must be quite cautious when approaching very high loading densities, to be certain that the regimes other than normal propellant regression are avoided.

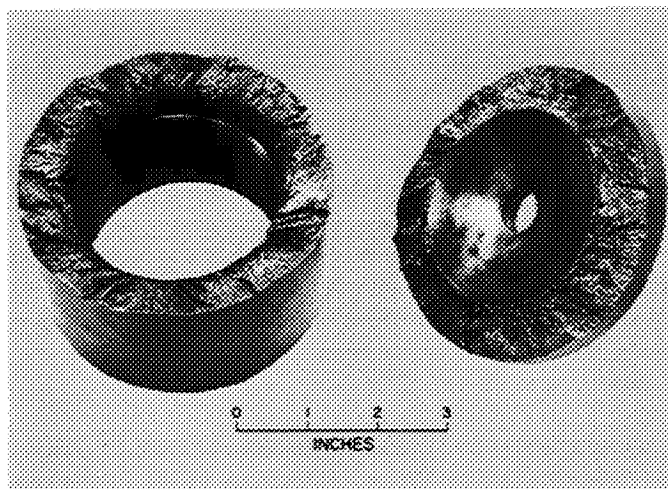


Fig. 14. End-cap subsequent to detonation

XI. Polymer Research

PROPULSION DIVISION

A. Ethylene Oxide-Freon 12 Decontamination Procedure: Reactions in the Decontamination Chamber and Effective Air-Flush Periods,

S. H. Kalfayan and R. H. Silver

1. Introduction

In connection with the ethylene oxide-Freon 12 decontamination procedures and practices, several problems as applied particularly to the sterilization of polymeric materials are being examined. Under consideration are: the quantitative estimation of ethylene oxide concentration in the decontamination chamber; the evaluation of various types of humidity sensors; the appropriate duration of vacuum and air-flush periods to remove the sterilant gas mixture from the chamber and from the polymeric materials exposed; and the chemical interactions that might take place between the gases present, i.e., ethylene oxide, water, Freon 12, and oxygen (from air).

This article summarizes the results obtained to date in the investigation of the reactions that could take place in the chamber, and the results from experiments intended to establish more effective air-flush periods.

2. Reactions in the Decontamination Chamber

The ethylene oxide decontamination procedure according to specification¹ is carried out at 50°C and 50%

relative humidity for six cycles of 28 h each. It was suspected that under these conditions, ethylene oxide might react with moisture, forming ethylene glycol, and also that Freon 12 (dichlorodifluoromethane) could hydrolyze to form HCl, especially in the presence of acidic or basic impurities originating from the polymeric materials.

To test this possibility, five 250-ml capacity ampules were filled with the chamber gases. Three of the ampules, containing only the chamber gases, a brass strip, and strips of polymeric materials, respectively, were subjected to 50°C for 180 h. The other two ampules, one containing a brass strip and the other strips of polymeric materials, were subjected to 58°C for 180 h.

After the heating period, analysis by mass spectrography showed a peak at mass 36 in all five ampules. The mass 36 peak, considered to be HCl, amounted to 0.06 mole % in the case of ampules subjected to 50°C, and 0.08 and 0.14 mole % in the case of ampules subjected to 58°C. The higher mole percent was obtained from the ampule containing the brass strip.

Only a trace of mass 36 peak was shown with the ethylene oxide-Freon 12 mixture sampled directly from the original gas cylinder.

It was concluded that HCl was formed in small amounts under the conditions of the decontamination process. This experiment indicated that the presence of polymeric products did not influence the formation of HCl, but that the presence of brass and increased temperature increased the rate of HCl formation.

¹"Environmental Specification, *Voyager* Capsule Flight Equipment, Type-Approval and Flight Acceptance Test Procedures for the Heat Sterilization and Ethylene Oxide Decontamination Environments," JPL Spec. VOL-50503-ETS.

Infrared spectrographic analysis of droplets obtained from one of the ampules showed the presence of ethylene glycol.

3. Establishing Optimum Air-Flushing Periods

At the end of each ethylene oxide-Freon 12 exposure cycle, the chamber is evacuated and flushed with ambient air for $2\frac{1}{2}$ to $3\frac{1}{4}$ h in order to free it and the decontaminated materials from sterilant gases.

There was strong indication from experience with the ethylene oxide decontamination of polymeric products that flushing for $2\frac{1}{2}$ to 3 h with air still left considerable amounts of absorbed sterilant gas mixture in the polymer samples. It was necessary to establish a more effective flushing period, since shortly after the ethylene oxide decontamination, samples were exposed to dry heat sterilization. The absorbed gases could damage the polymeric products at the higher temperature, before they had a chance to be desorbed.

The experimental approach consisted of the following: five uniform-sized strips were cut from each of five representative polymeric products (Figs. 1 and 2) and weighed.

They were then exposed to one cycle of ethylene oxide-Freon 12 decontamination. A sample strip from each product was taken out of the chamber before any flushing, and placed in separate ampules and sealed. This process was repeated after 2, 4, 6, and 8 h of flushing. The weight of the unflushed samples, and those flushed for 8 h were measured. The ampules were left for 24 h to provide enough time for desorption before the gases were analyzed by gas chromatography. Only Freon 12 and ethylene oxide were considered.

The results of this experiment are plotted in Figs. 1 and 2. The values of the peak areas for Freon 12 (Fig. 1) and ethylene oxide (Fig. 2) obtained for unflushed samples (zero hour) were set to equal 100%. Thus, the peak area obtained for Freon 12 from the analysis of the ampule containing the Stycast epoxy product (Fig. 1) after 2 h of flushing, was 80% of the area obtained from the ampule containing the unflushed sample. Failure to detect any Freon 12 or ethylene oxide in any of the ampules did not necessarily mean that the sample strip was freed of these gases after a period of flushing. With the exception of the glass-filled phenolic compound, all other four compounds still showed more weight after 8 h of flushing than they did before exposure to the decontamination process,

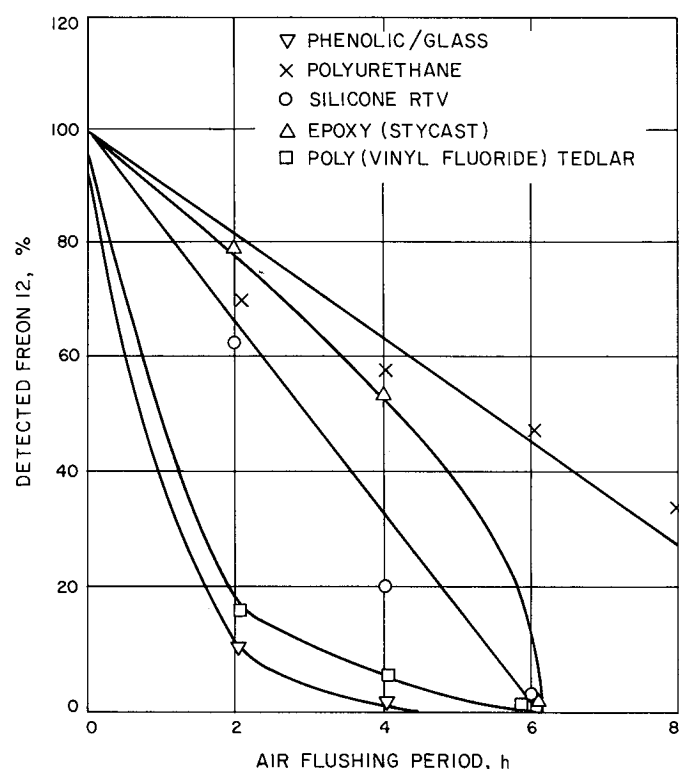


Fig. 1. Relationship of Freon 12 desorption with air-flush time

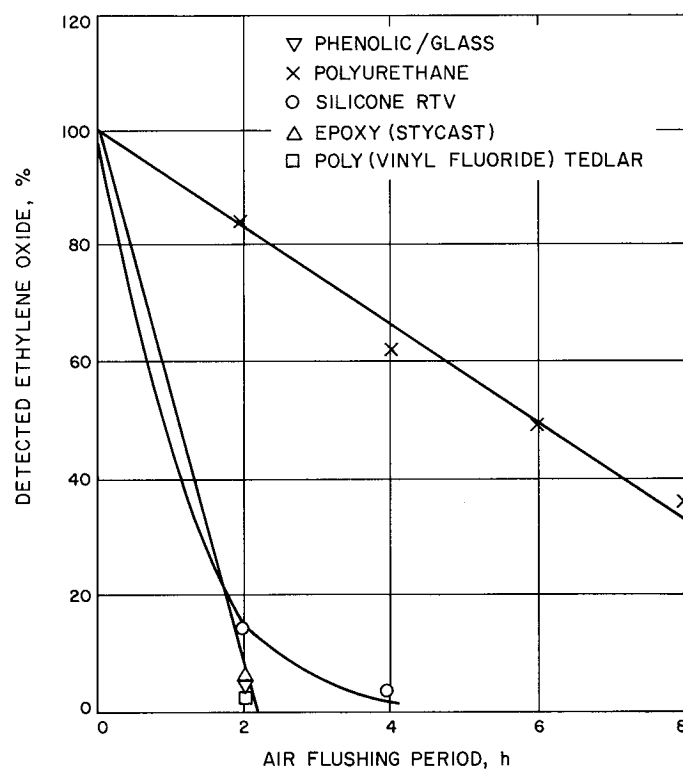


Fig. 2. Relationship of ETO desorption with air-flush time

although no gases could be detected after this period for three of them—the RTV silicone, the Stycast epoxy, and Tedlar. Weight measurements showed that these three compounds still retained, after 8 h of flushing, 5 to 10% of the total weight gained. The polyurethane product retained about 40% of its gained weight after 8 h of flushing.

The conclusion reached from these experiments is that flush time should be extended. Complete desorption at ambient pressure from polymeric materials may take days; however, the results of these experiments indicate that 90 to 100% of the absorbed sterilant gases could be expelled from most of the materials tested after 8 h of air flushing.

B. Thermally Stable Urethane Elastomers,

E. F. Cuddihy and J. Moacanin

1. Introduction

Recently, a new liquid diol prepolymer Telagen S (General Tire and Rubber Company) has become available. Its structure offers promise for the development of sterilizable elastomers. This prepolymer, a hydroxyl-terminated saturated polybutadiene, can be reacted with a diisocyanate and a triol to form a urethane elastomer. For this study an elastomer designated DU-1 (formulation parameters: $\text{NCO}/\text{OH}_{\text{total}} = 1.05$; $\text{OH}_{\text{triol}}/\text{OH}_{\text{total}} = 0.19$) was prepared from Telagen S (Batch No. 173H; equiv wt = 1040), 1,1,1-trimethylol propane triol, and an 80/20 mixture of 2,4 and 2,6 tolylene diisocyanate.

2. Analysis and Tests

Analysis has shown that Telagen S contains considerable proportions of preparative agents such as catalyst, as well as large quantities of monofunctional and nonfunctional polymer chains. In order to assess the effect of these materials on the properties and heat stability of the cured elastomer, a sample of DU-1 elastomer was extracted exhaustively with benzene. This procedure presumably removed the undesirable agents and any unreacted or nonfunctional Telagen S. Thirty weight percent of the initial DU-1 elastomer was removed as a soluble phase, and the resultant extracted elastomer was designated DU-1E.

The thermal stability of DU-1E was studied by following the changes in modulus of samples heated at 135°C in air, while that of DU-1 was studied by following the changes in modulus of samples heated also at 135°C in pure oxygen, pure nitrogen, air, and vacuum (1×10^{-6} torr).

The samples tested under pure oxygen and nitrogen were heated for only 96 h, while all the other samples were heated for a maximum exposure time of 318 h, which corresponds to the heat sterilization procedure of six consecutive 53-h cycles at 135°C. The results are shown in Fig. 3.

3. Results

Comparing first the unsterilized DU-1 and DU-1E elastomers, it is seen that DU-1E has a higher modulus than DU-1. The difference in the rubbery zone is identically that to be expected from the 30 wt % of soluble material.

For DU-1E, the modulus is observed to slightly improve after heating in air for 318 h at 135°C. The thermal stability of DU-1E elastomer is excellent when tested under this present heat sterilization procedure.

For DU-1, reasonable thermal stability in air is seen even though there is a noticeable drop in modulus after 318 h. Under vacuum however, DU-1 has significantly degraded in 318 h as attested by its modulus behavior. Comparing the test results after 96 h for DU-1 heated in oxygen, nitrogen, and vacuum reveals that the elastomer experienced the least reduction of modulus in oxygen and the greatest reduction when heated in vacuum, with an intermediate loss for the nitrogen exposure. The resulting spread in modulus values increases with increasing heating time, i.e., DU-1 degrades at a faster rate when heated under vacuum than when heated in air (oxygen).

These latter results are clearly consistent with known degradation and crosslinking mechanisms. The Telagen S backbone of DU-1 was prepared from polybutadiene which was hydrogenated to obtain a saturated system. However, the possibility of some residual unsaturation definitely remains, and these sites are susceptible to oxidative crosslinking, which would tend to increase the modulus. On the other hand, the curing of urethane elastomers results in the formation of certain thermally unstable linkages such as allophanates, biurets, or di-substituted ureas (Ref. 1). The dissociation of these linkages at elevated temperatures would tend to decrease the modulus. Furthermore, unpublished NMR studies² have indicated that an ether linkage can be incorporated into the Telagen S backbone during preparation. At elevated temperatures and in an acidic environment, these ether linkages can dissociate; analyses show Telagen S to be slightly acidic.

²D. D. Lawson, Polymer Research Sect.

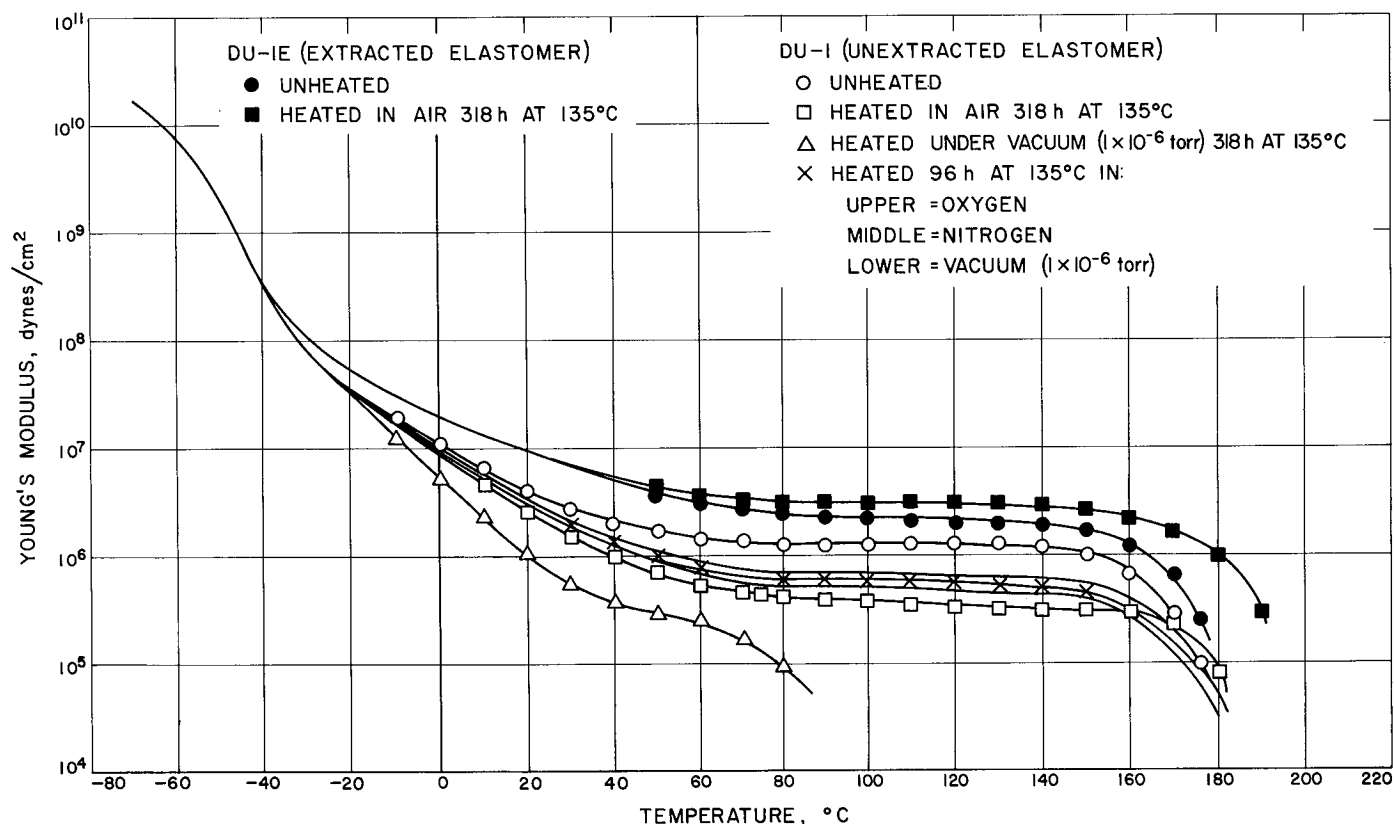


Fig. 3. Effect of thermal exposure on Telagen 5 containing urethane elastomers

In vacuum, where no oxidative crosslinking occurs, the elastomer degrades faster than in the presence of oxygen, where compensating crosslinking can occur to preserve the level of the modulus.

For DU-1E heated in air, the modulus slightly improved while the modulus of DU-1 heated in air decreased, thus suggesting that the extraction process removed constituents which either promoted the dissociation of thermally unstable linkages (i.e., an acidic ingredient) or decreased the extent of oxidative crosslinking (i.e., an antioxidant).

Reference

1. Tobolsky, A. V., *Properties and Structures of Polymers*, Chap. 5, John Wiley and Sons, Inc., New York, 1960.

C. A Relationship Between Maximum Packing of Particles and Particle Size, R. F. Fedors

1. Introduction

Experimental studies have demonstrated that both the viscosity for uncrosslinked fluids and the small strain modulus for crosslinked fluids which contain spheroidal

particulate fillers can be expressed by an equation of the form (Ref. 1):

$$\frac{\eta}{\eta_0} = \left(\frac{\phi_{\max}}{\phi_{\max} - \phi} \right)^{2.5} \quad (\text{uncrosslinked}) \quad (1)$$

or

$$\frac{E}{E_0} = \left(\frac{\phi_{\max}}{\phi_{\max} - \phi} \right)^{2.5} \quad (\text{crosslinked}) \quad (2)$$

where η and η_0 are the viscosities of the filled and unfilled uncrosslinked fluid, respectively; E and E_0 are the moduli of the filled and unfilled crosslinked fluid, respectively; ϕ is the volume fraction of filler present; and ϕ_{\max} is the maximum volume fraction of filler which the fluid can accept and still maintain a two-phase system. Thus, a knowledge of the values of only two parameters, η_0 (or E_0) and ϕ_{\max} , permits one to estimate the viscosity (or modulus) of a system as a function of ϕ . In practice, η_0 (or E_0) is easily measured while ϕ_{\max} , on the other hand, is usually more difficult to determine.

For relatively large-size particles, the particles are poured into a container and the ratio of the volume of the particles to the volume they occupy is determined. This ratio is defined as ϕ_{\max} . In this kind of experiment, it is known that the measured ϕ_{\max} value depends on the container size (i.e., on the number of particles present) and shape, and on the imposition of mechanical vibration to the system either during or after the loading of the particles. Thus, Scott (Ref. 2), working with 2×10^4 uniform steel balls, $\frac{1}{8}$ in. in diameter, and using a variety of container types, found that without mechanical vibration, a loose-random packing is obtained with $\phi_{\max} = 0.60$; when mechanical vibration is employed, however, dense-random packing occurs and $\phi_{\max} = 0.63$ is obtained. These ϕ_{\max} values were approximately independent of container geometry. Susskind et al. (Ref. 3), working with up to 1.5×10^5 uniform glass beads 0.118 in. in diameter and with $\frac{1}{8}$ -in. steel balls, also with a variety of container geometries, obtained $\phi_{\max} = 0.63$ for the loose-random-packing and $\phi_{\max} = 0.65$ for the dense-random packing.

For small-size particles, a method commonly employed involves determining the ratio of the volume of the particles to the volume of the packed bed formed when a slurry containing the particles is permitted to sediment. For very small particles, high-speed centrifugation is usually used to speed up sedimentation of the slurry (Ref. 1). Using this technique on glass beads (diam = 53μ) in mineral oil and on aluminum (diam = 11μ) in both mineral oil and low molecular weight polypropylene glycol, it was found that the calculated ϕ_{\max} increased as either ϕ of the slurry and/or the volume of slurry employed was increased (Ref. 1).

In addition to these effects, it is also known that ϕ_{\max} decreases as the particle dimensions are decreased for small-size particles. For example, with copper in mineral oil, $\phi_{\max} = 0.64$ for particles with an average diameter of 55μ and $\phi_{\max} = 0.347$ when the average diameter is 12μ (Ref. 1).

It has also been established that the particle-size distribution has an effect on the observed ϕ_{\max} value. Westman and Hugill (Ref. 4) observed that for a mixture or blend of uniform particles of two or more sizes, a value of ϕ_{\max} which was higher than the ϕ_{\max} of any of the components was obtained at a given blend ratio. Both the magnitude of the largest ϕ_{\max} value and the blend ratio at which it occurred depended on the diameter ratios. For example, they reported $\phi_{\max} = 0.992$ for a blend of five components, each of which was uniform in size with a $\phi_{\max} = 0.62$.

Thus, any attempt to predict a ϕ_{\max} value from other more easily measured parameters must consider three

factors: (1) calculation of ϕ_{\max} for random packing of uniform particles, (2) variation of ϕ_{\max} with particle size, and (3) variation of ϕ_{\max} with particle-size distribution. Only the first two factors will be discussed in this report. In what follows, theoretical attempts to calculate ϕ_{\max} for random packing will first be discussed. Then, the existing theory for the effect of particle size on ϕ_{\max} will be described, and finally an alternative explanation for this effect will be presented.

2. Discussion

a. Packing of uniform spheres. Rogers (Ref. 5) has been able to show from theoretical considerations that, for the packing of uniform spheres, ϕ_{\max} cannot exceed the value of 0.7797... for any packing whatever. Since the largest ϕ_{\max} known for an ordered packing, that of the hexagonal close packing is 0.7404..., Coxeter has postulated that a random packing may exist which has a ϕ_{\max} greater than that of the hexagonal close-pack structure (Ref. 6). Rogers has also shown that in two dimensions, i.e., the packing of uniform circles in two-dimensional space, the two-dimensional analog of ϕ_{\max} , χ_{\max} , cannot exceed the value of 0.9069... which corresponds to the value for an ordered hexagonal array of circles. For the one-dimensional case, i.e., the packing of uniform line segments onto an interval, the one-dimensional analog of ϕ_{\max} , ψ_{\max} , cannot exceed unity. Figure 4 shows the intriguing (nonlinear) dependence of maximum packing on the dimensionality of space. Rogers has been able to extend these results to higher dimensionalities n , and the result, valid only for large n , is

$$\phi_{\max} \leq \frac{n}{e} \left(\frac{1}{\sqrt{2}} \right)^n \left[1 + O\left(\frac{1}{n}\right) \right] \quad (3)$$

where e is the base of the natural logarithm and $O(1/n)$ is a monotonically decreasing function. Equation (3) predicts that higher dimensional space must be essentially empty, since $\phi_{\max} \rightarrow 0$ as $n \rightarrow \infty$. These results of Rogers establish an upper limit to the value of ϕ_{\max} independent of the mode of packing. We now consider the estimation of ϕ_{\max} for a particular mode of packing, i.e., random packing.

For the one-dimensional case, this problem has been rigorously solved under the heading of the "parking problem." Consider the following process in which cars of unit length are parked on a street whose length is greater than unity: the first car is parked at random; if additional space remains, a second car is parked at random, also. This process continues until no empty interval or space remains which is large enough to accommodate another car. Renyi (Ref. 7) has shown that the average number of parked cars

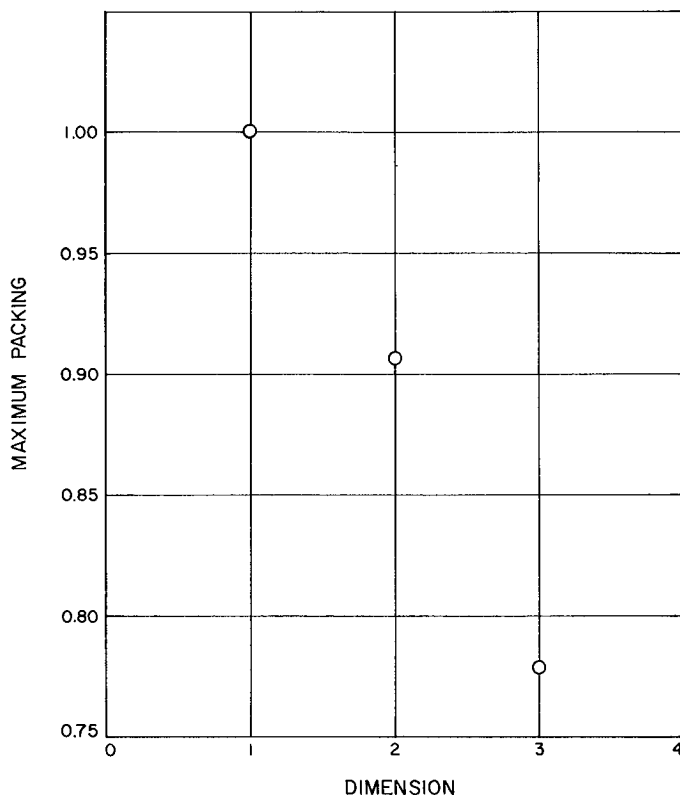


Fig. 4. Dependence of maximum packing on dimension of space

N_x on a street of length x is

$$N_x = cx + (c-1) + O\left(\frac{1}{x^n}\right) \quad (4)$$

where c is a constant equal to 0.7479... For large x , i.e., for a street long compared to the car length, the line fraction occupied by the cars, ψ_{\max} , is

$$\psi_{\max} = \lim_{x \rightarrow \infty} \frac{N_x}{x} = c = 0.7479 \quad (5)$$

It is very tempting to relate ψ_{\max} for the one-dimensional case to ϕ_{\max} in the three-dimensional case. However, since the real case concerns uniform spheres, it is evident that a line passing through randomly packed particles would necessarily not intersect all the particles diametrically; some particles will be intersected along cords the length of which will be determined by a probability distribution function, i.e., the length of the car in the parking problem is variable rather than a constant, as was assumed in the derivation of Eq. (4).

Another objection to the straightforward application of Eq. (4) was pointed out by Greet (Ref. 8) who was able to derive the actual density distribution function for the parking problem, rather than just the average value expression given in Eq. (4). His result is

$$P(\psi_{\max}) = \frac{B}{A} \exp \beta \left[\frac{(1-\psi_{\max})}{\psi_{\max}} \right] \quad (6)$$

where B is a normalization constant and A and β are functions of ψ_{\max} . This density distribution function gives a most probable value of ψ_{\max} equal to 0.667... which differs from the average value provided by Eq. (4) of 0.749... Greet has pointed out that the most probable value of an observed quantity must also be the average value, and this is not the case in one dimension. In addition, the distribution function is relatively broad, and there is a finite probability of finding all densities between $\frac{1}{2}$ and 1, i.e., between a street half full of cars and one completely full.

For $n \geq 2$, no strict probabilistic analysis of the packing problem appears to have been published. However, in attempting to extend the parking problem results to higher dimensions, Palasti (Ref. 9) conjectured that in two dimensions, i.e., when a rectangle is filled at random by unit squares, the average filled area fraction is $c^2 = (0.7479...)^2 \cong 0.56...$. In experiments reported in the paper on the random filling of rectangles with unit squares, the observed χ_{\max} value was 0.56, which agrees with the predicted value. Palasti then extends the conjecture to n dimensions and suggests that, in general, the average filled fracture of space is simply c^n . For the case of $n = 3$, the average value of ϕ_{\max} for the random packing of non-overlapping unit cubes into a parallelepiped is $c^3 \cong 0.42$. Since spheres can be packed more densely without overlap than can cubes, Palasti's average ϕ_{\max} value should be considered as a minimum estimate for the case of the packing of uniform spheres.

An attempt has also been made by Landel and Moser to estimate ϕ_{\max} for the random packing of uniform spheres (Ref. 10). In their treatment, it is assumed that ϕ_{\max} for the random packing of spheres will be intermediate in value between the ϕ of the most dense ordered packing, i.e., hexagonal close packing with $\phi = 0.74$ and the ϕ characteristic of the least dense, but stable, ordered packing which is assumed to be given by the simple cubic array with $\phi = 0.52$. It should be noted, however, that a packing with a ϕ value of 0.125 has been reported (Ref. 11). To find the most probable or average array, each particle contact is considered to occupy a spherical sector. Thus, in the

hexagonal close-pack array with a coordination number of twelve, each particle is considered to be composed of twelve independent sectors. For the most stable open structure, the coordinated number is six, and hence each particle is composed of six sectors. If all sectors have an equal probability of being filled or remaining empty after the required minimum of six have been filled, the probability of having n_f filled sectors out of a total of n sectors is taken as

$$P(n_f) = \frac{n!}{2n! (n - n_f)!} \approx \frac{1}{2} \left[\frac{1}{(1 - x)} \left(\frac{1}{x} - 1 \right)^x \right]^n \quad (7)$$

where x is the fraction of filled sectors. This last expression in terms of x is valid only for large n . The most probable value of x is $\frac{1}{2}$. Hence the most probable value for $\phi_{\max} = (0.74 - 0.52)/2 + 0.52 = 0.63$, a value in excellent agreement with the experimentally measured ϕ_{\max} .

A drawback of this treatment is the assumption that the sectors can be taken as independent of one another. This is not the case, since the sectors' association with a given particle is not independent. To look at the situation in a different light, consider the process where particles are removed from the hexagonal close-packed array one at a time until the ϕ value reads 0.63. For the first few particles, 12 sectors will be removed each time a particle is withdrawn. However, as more and more particles are withdrawn, it is clear that the number of sectors removed per particle will decrease because the structure contains fewer particles. The use of Eq. (7) is equivalent to the assumption that the number of sectors removed per particle is independent of stage at which the particle is withdrawn.

In addition to deriving a value for ϕ_{\max} for relatively large particles, they also introduce the concept of a lower limit on ϕ_{\max} which is envisaged as occurring in the limiting case for uniform spheres where the particle-particle interaction is very great. It is conjectured that as the interaction increases, the minimum value for ϕ_{\max} approaches zero. Using the same arguments as before, i.e., Eq. (7), the most probable state occurs when $x = \frac{1}{2}$, i.e., half the sectors are filled. Since interaction between particles is not expected to change the maximum value for ϕ , i.e., that for the hexagonal close pack, the most probable value for the lower limit on $\phi_{\max} = (0.74 - 0)/2 + 0 = 0.37$.

This development suffers the same shortcomings mentioned above, i.e., the assumption that the particle sectors

are independent. Further, there does not seem to be any experimental data to support the existence of a lower limit on ϕ_{\max} .

b. Dependence of packing on particle radius

Relationship based on particle-particle interaction. Using the sector method, Landel and Moser also derive an expression for the dependence of ϕ_{\max} on particle size (Ref. 10). A randomly packed bed of uniform spheres which are subject to particle-to-particle interaction is divided into arbitrarily small volumes called cells. The probability of the i th cell to have n_{fi} filled sectors out of a total number n_i is taken as

$$W_i = \frac{n_i!}{2n_{fi}! (n_i - n_{fi})!} \quad (8)$$

After summing over all i , the resulting equation is

$$\frac{n_{fi}}{n_i} = \frac{1}{1 + A \exp(\alpha S_i)} \quad (9)$$

where A and α are undetermined multipliers, and S_i is the interaction energy in the i th cell. Equation (9) is transformed into

$$\phi_{\max} = 0.37 \left[1 + \frac{1}{1 + 0.424 \exp \beta/2r} \right] \quad (10)$$

where β is a parameter related to the surface energy of a particle and r is the particle radius. In the derivation of Eq. (10) the sectors, as before, were assumed to be independent of one another.

It was further proposed on the basis of experimental data, that β could be related to the constant C of the Brunauer-Emmett-Teller theory for the adsorption of gases on solids by means of the following equation (Ref. 10):

$$\log \beta = 2.7(\log C - 1.28) \quad (11)$$

where the C values are determined using krypton as the adsorbent. Thus, according to Eqs. (1), (10), and (11), a knowledge of the surface energy of a particle, i.e., the C value, and the particle size permits one to estimate ϕ_{\max} and, hence, the viscosity of slurries.

Equation (10) is shown in graphical form in Fig. 5. The curve is S-shaped with an upper bound ϕ_{\max} equal to 0.63, which is reached according to Eq. (10) when $\beta/2r \rightarrow 0$, i.e., when either $\beta \rightarrow 0$ or $r \rightarrow \infty$. A lower bound ϕ_{\max}

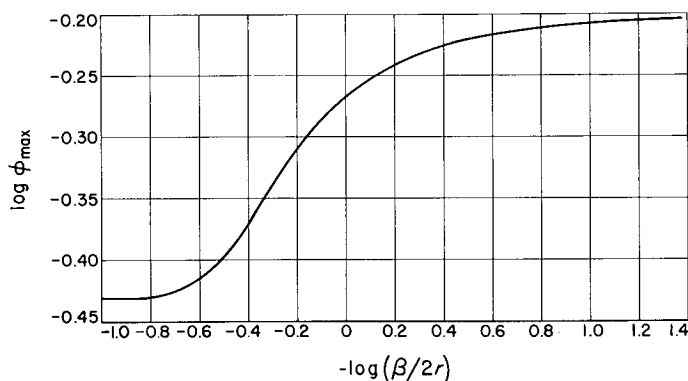


Fig. 5. Variation of $\log \phi_{\max}$ with $\log \beta/2r$ as predicted by Eq. (10)

equal to 0.37 occurs when the ratio $\beta/2r \rightarrow \infty$, i.e., when either $\beta \rightarrow \infty$ or $r \rightarrow 0$. The parameter β is, in the theory, related to the degree of adherence of the particles. For particles with little or no tendency to adhere, ϕ_{\max} approaches the upper bound value of 0.63 independent of the precise value of the particle size. On the other hand, for particles with a great tendency to adhere, ϕ_{\max} approaches the lower bound value of 0.37. However, it is likely that Eq. (10) would break down before any lower bound, if it exists, were reached. For as ϕ_{\max} decreases, the theory would predict that particle-particle interaction increases, since the lower bound will be approached only for very large β values. It can be argued, however, that if this were true, then the agglomerates or clusters produced because of interactions between particles must be very highly asymmetrical in shape (e.g., rodlike) since, if the agglomerates were symmetrical, these would merely correspond to particles of larger size, and the ϕ_{\max} value would be expected to be high, i.e., to approach the upper bound value of 0.63. For highly asymmetrical structures or clusters, the concept of a unique ϕ_{\max} itself as well as the validity of Eqs. (1) and (2) should be seriously questioned.

Further, if agglomerates are symmetrical in form, it is reasonable to expect that they will vary in size, which would imply a distribution of particle or agglomerate sizes. The existence of a distribution of sizes will require changes in the value of the upper bound in Eq. (10) and perhaps in the form of the equation as well, since its derivation assumed uniform particles.

In addition, it is surprising that a measure of particle-particle interaction by itself, and without regard to particle-fluid interaction, as required by Eq. (10) is a sufficient indication of the ability of particles to agglomerate in a slurry.

3. Relationship Based on Particle-Fluid Interaction

An alternative approach is based on the premise that particles immobilize a fraction of the fluid, which leads to an increase in the effective volume occupied by the particles.

Consider a group of particles of radius r_i , $i = 1, 2, \dots, N$, where N is the total number of particles present. Assume that each particle adsorbs a surface layer of fluid so that now the radius of the i th particle becomes $r_i + \delta_i$, where δ_i is the thickness of the layer. The volume of the particle itself is $(4\pi/3)r_i^3$, while the volume of the particle plus layer is $(4\pi/3)(r_i + \delta_i)^3$. The basic assumption is now made that the value of ϕ_{\max} calculated when the surface layer is neglected is related to the true value $\phi_{\max, T}$ by means of the following equation

$$\frac{\phi_{\max}}{\phi_{\max, T}} \equiv \frac{\sum_{i=1}^N n_i r_i^3}{\sum_{i=1}^N n_i r_i^3 \left(1 + \frac{\delta_i}{r_i}\right)^3} \quad (12)$$

where n_i are the number of particles of radius r_i . In order to estimate δ_i we make the additional assumption that the ratio of the volume of fluid adsorbed by a particle to the surface area of the particle is a constant. This assumption leads to the relationship

$$\left(1 + \frac{\delta_i}{r_i}\right)^3 = 1 + \frac{3k}{r_i} \quad (13)$$

where k is the constant of proportionality between volume adsorbed and particle surface area. Substituting this result in Eq. (12), there is obtained:

$$\phi_{\max} = \phi_{\max, T} \left[\frac{1}{1 + 3k \left(\frac{\sum_{i=1}^N n_i r_i^2}{\sum_{i=1}^N n_i r_i^3} \right)} \right] \quad (14)$$

If a monodisperse system is considered, Eq. (14) reduces to the simple result

$$\phi_{\max} = \phi_{\max, T} \left(\frac{1}{1 + (3k/r)} \right) \quad (15)$$

It is not, however, necessary to restrict Eq. (14) to the case of uniform particles. As a matter of fact, it will now be shown that the form given by Eq. (15) is invariant to the type of particle-size distribution one considers, i.e.,

this form is valid for any distribution of sizes. It is evident that the radius of the i th particle can be expressed as some multiple a_i of any other particle dimension, such as the number average particle radius r , e.g.,

$$r_i = a_i r \quad (16)$$

The a_i s depend on the distribution of sizes and also on the choice of the reference, which in Eq. (16) is \bar{r} . The summations occurring in Eq. (14) become

$$\frac{\sum_{i=1}^N n_i r_i^2}{\sum_{i=1}^N n_i r_i^3} = \frac{\sum_{i=1}^N n_i a_i^2 r^2}{\sum_{i=1}^N n_i a_i^3 r^3} = \frac{1}{r} \frac{\sum_{i=1}^N n_i a_i^2}{\sum_{i=1}^N n_i a_i^3} = \frac{C}{\bar{r}} \quad (17)$$

where C is a constant whose magnitude depends on both the type of distribution and the reference size. Provided the same reference size is used, e.g., r , and provided the nature of the size distribution does not vary with particle size or is only a weak function of particle size, Eq. (15) can be written more generally as

$$\phi_{\max} = \phi_{\max, T} \frac{1}{(1 + k'/r)} \quad (18)$$

where k' is $3kC$ and \bar{r} is any convenient average of reference size. As an example, consider the simple case of the most probable or random distribution defined by

$$\frac{n_i}{N} = \frac{1}{\bar{r}} \exp\left(-\frac{r_i}{\bar{r}}\right) \quad (19)$$

If the summations in Eq. (17) are replaced by integrations, there is obtained:

$$\frac{\sum_{i=1}^N n_i r_i^2}{\sum_{i=1}^N n_i r_i^3} = \frac{\int_0^\infty r_i^2 \exp\left(-\frac{r_i}{\bar{r}}\right) dr_i}{\int_0^\infty r_i^3 \exp\left(-\frac{r_i}{\bar{r}}\right) dr_i} = \frac{1}{3\bar{r}} \quad (20)$$

Thus, for this particular distribution of sizes, $C = 1/3$ when the reference size is \bar{r} .

Figure 6 shows the form of Eq. (18) when $\log \phi_{\max}$ is plotted against $-\log \bar{r}$. In order to employ Eq. (18), experimental data for the variations of ϕ_{\max} with particle size are plotted in log-log coordinates and superposed on the master curve in Fig. 6 to obtain the best fit. The vertical shift is $\log \phi_{\max, T}$, while the horizontal shift corresponds to $\log k'$. Note that the vertical and horizontal shifts are

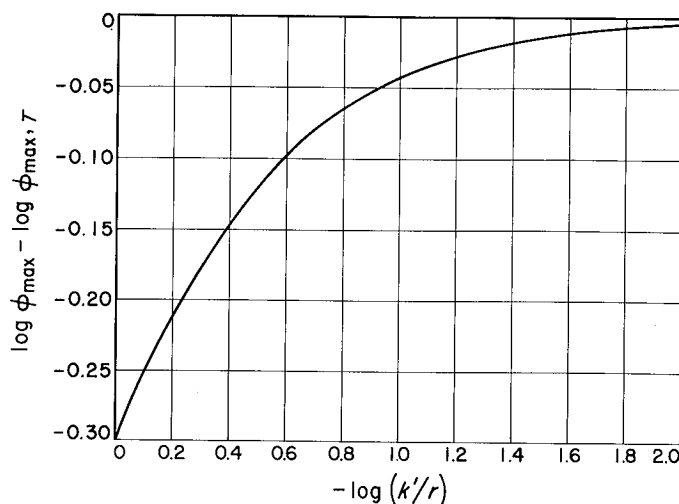


Fig. 6. Variation of $\log \phi_{\max}$ with $\log k'/r$ as predicted by Eq. (18)

independent so that an unambiguous estimate for $\phi_{\max, T}$ is obtained even when the form of the particle-size distribution is not known.

As an indirect test of the form of Eq. (18), the data of Zettlemoyer and Lower on the viscosity of sized (nearly monodisperse) CaCO_3 in polybutene will be employed (Ref. 12). The value of ϕ_{\max} for each of the CaCO_3 particle sizes used was not directly measured. However, using their viscosity data in conjunction with Eq. (1), the ϕ_{\max} values shown in Table 2 were estimated. Equation (1) was used in the form $(\eta_0/\eta)^{1/4} = 1 - (\phi/\phi_{\max})$, and plots of the left-hand side against ϕ were prepared for each particle size. ϕ_{\max} was then obtained from the least-square estimate of the slope.

These data are shown in Fig. 7 as the circles and the master curve according to Eq. (18) is shown as the full curve. From the shifts required for fit, $\phi_{\max, T} = 0.634$ and $k = 140 \text{ \AA}$, assuming the fractions are monodisperse. This k value corresponds to a thickness of the adsorbed liquid layer of 140 \AA , which is unreasonably high. The liquid used was polybutene with a molecular weight of 530

Table 2. ϕ_{\max} as a function of particle size for CaCO_3

Radius, μ	ϕ_{\max}
0.065	0.377
0.09	0.443
0.115	0.470
0.195	0.510

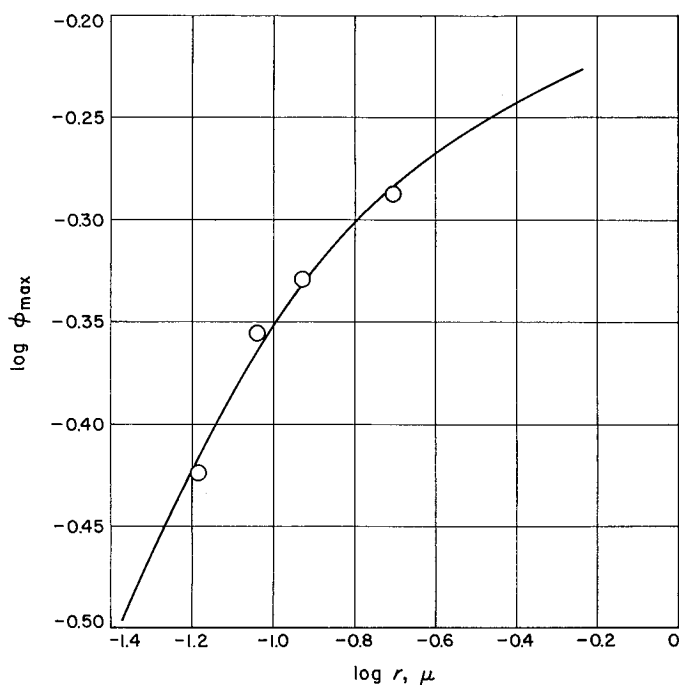


Fig. 7. Comparison of data on CaCO_3 -polybutene with the prediction of Eq. (18)

g/mole. The maximum thickness of an adsorbed layer one molecule thick would be only about 30 Å.

This discrepancy can be explained on the basis of particle agglomeration. When particles cluster or agglomerate, the interior regions of such clusters will contain entrained liquid which does not contribute to the flow of particles. If it is assumed that the volume of entrained liquid varies as the effective surface area of the cluster, then Eq. (18) remains valid; but now the k' value will contain a contribution from both the adsorbed surface layer and the entrained liquid,

$$k' = k'_s + k'_{en} \quad (21)$$

where k'_s is the contribution from the adsorbed layer and k'_{en} the contribution from the entrained liquid. It may be noted that for a close-packed array, the volume of entrained liquid would be expected to vary as the volume rather than as the surface area as assumed here. However, the behavior of a loosely packed array may differ from that of a closely packed structure.

To the extent that clustering occurs, then the number average radius of the agglomerates would be more appropriate to use than the radius of the monodisperse particle. This change in variable will not affect the esti-

mated value of $\phi_{\max, T}$; it will only modify somewhat the estimate for k' , and hence δ .

Figure 8 shows a comparison of the equation of Landel and Moser (the full curve) with Eq. (18) shown as the dashed curve. In obtaining this fit, it was assumed that $\phi_{\max, T}$ in Eq. (18) is 0.63. A better fit is obtained if $\phi_{\max, T}$ is taken as 0.645. The value of k' is related to β by the relation $k' = 11.5\beta$. For ϕ_{\max} values greater than about 0.4 (the two equations are in very good agreement and to within the usual scatter in the experimental data) the two equations may be taken as identical.

In order to study the reasons for the close correspondence between the two equations, we expand each in powers of $1/r$. Thus, Eq. (10) becomes to the fourth power in $1/r$.

$$\begin{aligned} \phi_{\max} = 0.63 \left[1 - \frac{a}{2(1+a)(2+a)} \frac{\beta}{r} \right. \\ - \frac{a(1-a)}{8(1+a)^2(2+a)} \frac{\beta^2}{r^2} - \frac{a(1-4a+a^2)}{48(1+a)^3(2+a)} \frac{\beta^3}{r^3} \\ \left. - \frac{a(1-11a+11a^2-a^3)}{284(1+a)^4(2+a)} \frac{\beta^4}{r^4} - \dots \right] \quad (22) \end{aligned}$$

where a is the preexponential factor, 0.424. The corresponding expression for Eq. (18) becomes

$$\phi_{\max} = 0.63 \left[1 - \frac{3k'}{r} + \frac{9k'^2}{r^2} - \frac{27k'^3}{r^3} + \frac{81k'^4}{r^4} - \dots \right] \quad (23)$$

Except for the positive sign of the term involving k'^2/r^2 in Eq. (23), the two expansions are formally equivalent up to at least the fourth power of the radius.

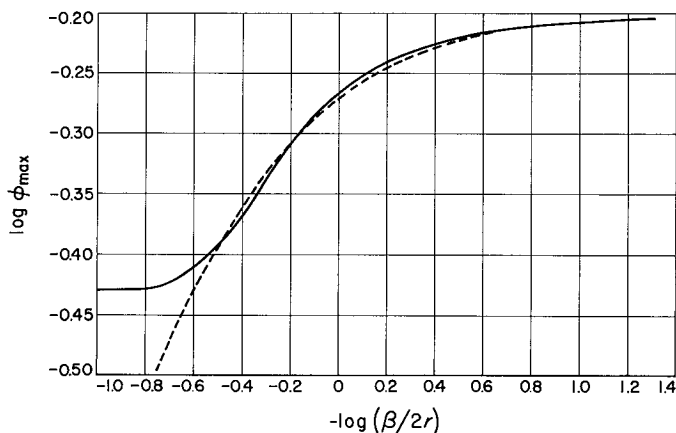


Fig. 8. Comparison of Eq. (10) with Eq. (18)

If Eq. (10) is considered to be empirical in nature, then using the same sort of argument involved in obtaining Eq. (18), it is easy to show that Eq. (10) can be also independent of the particle-size distribution.

4. Summary

A relationship between ϕ_{\max} and particle size has been developed. The basis for the variation of ϕ_{\max} with r is considered to depend primarily on the immobilization and entrainment of liquid on both the particle surface and also in the interior of clusters of particles.

References

1. Landel, R. F., Moser, B. G., and Bauman, A. J., *Proceedings of the Fourth International Congress on Rheology*, Brown University, Providence, Rhode Island, ed. E. H. Lee, Vol. II, p. 663, Interscience Publishers, New York, 1965.
2. Scott, G. D., *Nature*, Vol. 188, p. 908, 1960.
3. Susskind, H., Winsche, W. E., and Becker, W., Brookhaven National Laboratory, Report BNL 50022 (T-441), June 1966.
4. Westman, A. E. R., and Hugill, H. R., *J. Am. Ceram. Soc.*, Vol. 13, p. 767, 1938.
5. Rogers, C. A., *Proc. London Math. Soc.*, Vol. 3, p. 609, 1958.
6. Coxeter, H. S. M., *Illinois J. Math.*, Vol. 2, p. 746, 1958.
7. Renyi, A., *Publ. of the Math. Inst. Hungar. Acad. Sci.*, Vol. 3, p. 109, 1958.
8. Greet, R. J., *J. Appl. Phys.*, Vol. 37, p. 4377, 1966.
9. Palasti, I., *Publ. Math. Inst. Hungar. Acad. Sci.*, Vol. 5, p. 353, 1960.
10. Moser, B. G., and Landel, R. F., "Rheology of Slurries III: A Theory of the Sedimentation Volume for Systems Containing Uniform Spheres." Paper presented at the Pacific Conference on Chemistry and Spectroscopy, Anaheim, California, November 1967.
11. Scheidegger, A. E., *The Physics of Flow Through Porous Media*, Macmillan Company, New York, 1960, results of Hrubisch quoted on p. 20.
12. Zettlemoyer, A. C., and Lower, G. W., *J. Colloid Sci.*, Vol. 10, p. 29, 1955.

XII. Research and Advanced Concepts

PROPULSION DIVISION

A. Pressure Distribution Along the Wall of an Axisymmetric Second-Throat Diffuser for Ambient Temperature Air Flow, R. F. Cuffel, P. F. Massier, and L. H. Back

1. Introduction

Supersonic exhaust diffusers are frequently used during ground level tests of rocket engines to reduce the back pressure at the nozzle exit below that of the local atmosphere. A sufficient reduction in the back pressure will allow the engine tests to be conducted at the design chamber pressure without shock-induced flow separation occurring in the divergent portion of the nozzle. Numerous investigations have been conducted, including those of Massier and Roschke (Refs. 1 and 2), which pertain to the influence of configuration on diffuser performance. However, very little information exists on heat transfer, boundary layer structure, and wall-pressure distributions with sufficient spatial resolution to evaluate the flow field in an axisymmetric diffuser. The purpose of this investigation is to acquire knowledge in each of these areas which will be useful for understanding and establishing the cooling requirements of second-throat supersonic diffusers.

The present discussion, however, pertains only to wall-pressure distributions for a flow in which there was a negligible amount of heat transfer. Such adiabatic flow tests have been conducted at stagnation pressures for which the diffuser was started (full flowing nozzle) and at lower pressures for which the diffuser had not started (shock-induced flow separation in the nozzle). These

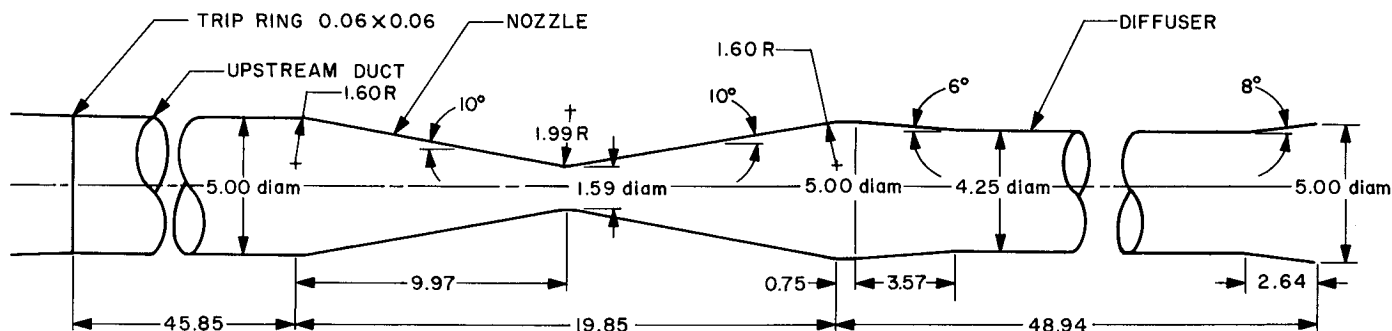
measurements indicate the extent of regions in which flow accelerations as well as decelerations can occur in such a diffuser. Baker and Martin (Ref. 3) have presented comparatively detailed wall-pressure distributions for a constant-diameter diffuser with a diameter larger than that of the nozzle exit.

2. Experimental Apparatus

A schematic drawing of the experimental apparatus is shown in Fig. 1. The upstream flow conditioning system (not shown in this figure) is described in Ref. 4. Compressed air in which the boundary layer was turbulent entered the nozzle after flowing through a 5.0-in.-diam duct about 46 in. long. The flow then proceeded through the nozzle and diffuser and was discharged into the atmosphere. The wall static pressures in the diffuser were measured with transducers accurate to within 0.05 psi.

3. Results

The diffuser started at a stagnation pressure of 81.5 psia, which is well below the starting pressure of 109 psia, computed using plane one-dimensional flow theory (Ref. 1) for the diffuser inlet to nozzle throat area ratio of 9.9. There was negligible hysteresis, i.e., the minimum starting pressure was the same as the minimum operating pressure. Wall static pressure distributions along the diffuser at inlet stagnation pressures above and below both the minimum starting and operating pressures are shown in Fig. 2.



DIMENSIONS IN INCHES

Fig. 1. Experimental apparatus

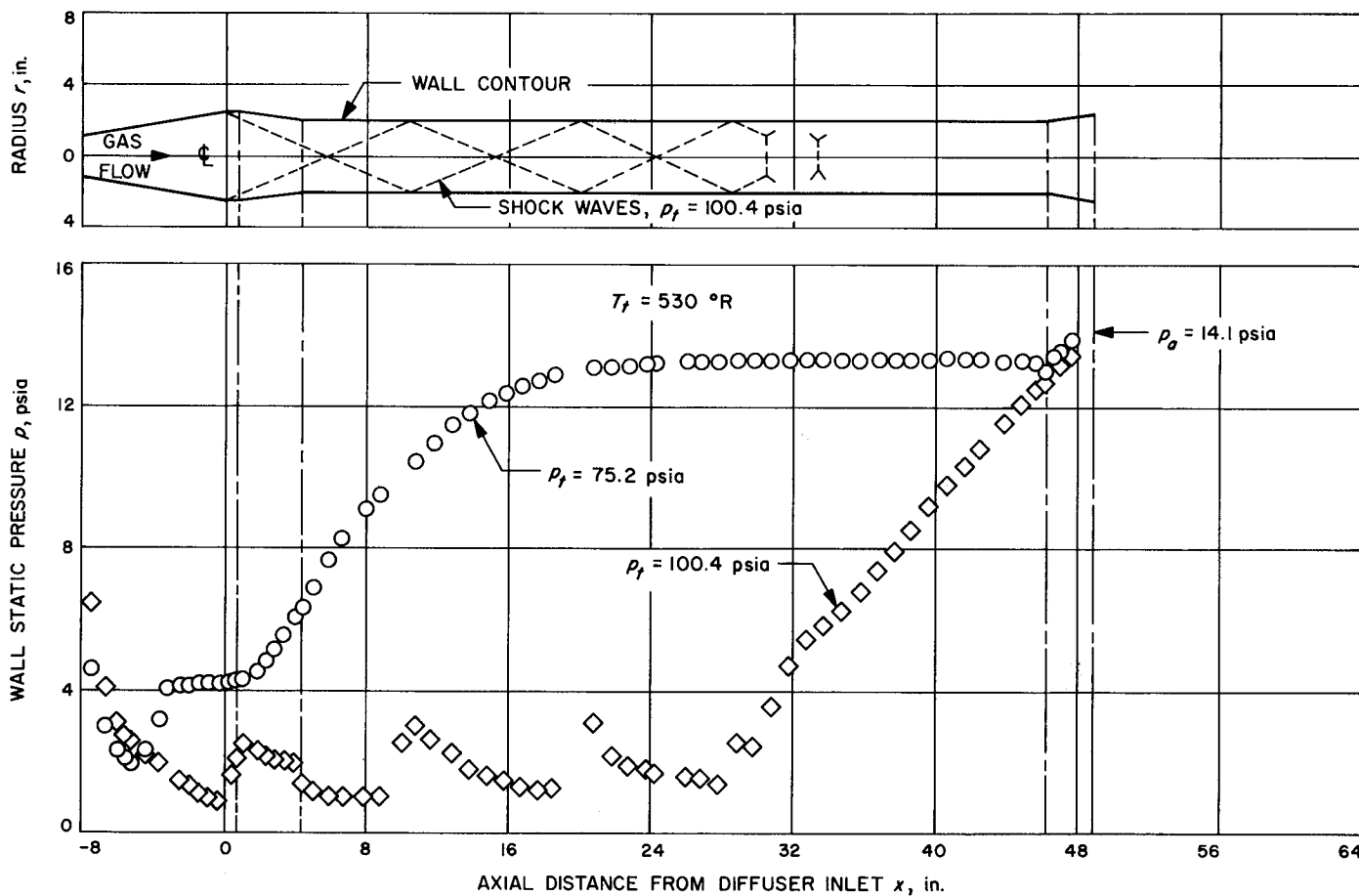


Fig. 2. Diffuser wall pressure distributions

At a stagnation pressure of 75 psia, which is slightly below the minimum starting pressure, the static pressure ratio across the diffuser p_e/p_a was about 0.30. The diffuser inlet pressure was the same as the nozzle exit pressure p_e , and the diffuser exit pressure was the same as atmospheric pressure p_a . For this condition the flow separation and an associated wall pressure rise occurred in the nozzle at a diameter of about 3.4 in. This is a smaller diameter than that of the diffuser throat. After separation in the nozzle, the flow must have subsequently reattached in the convergent part of the diffuser to undergo the deceleration from supersonic to subsonic flow indicated by the steady rise of the wall pressure in the upstream region of the diffuser.

At a stagnation pressure of 100 psia, which is above the minimum starting pressure, the nozzle flowed full with a Mach number of 3.85 at its exit, and p_e/p_a was about 0.064. In the upstream portion of the diffuser the flow was decelerated and accelerated several times, as evidenced by the repeated rise and fall of the wall static pressures in Fig. 2. Note that the axial distances over which the decelerations occurred are relatively short compared with the lengths of the acceleration regions. This behavior indicates the formation of an oblique shock at the entrance of the diffuser as the flow underwent a 16-deg compressive turning, and the subsequent reflections of the oblique shock from the centerline and wall. This shock structure is indicated by the dashed lines within the wall contour above the data in Fig. 2. The pattern is similar to those observed in Schlieren pictures of flow through parallel plate diffusers (e.g., Ref. 5). Downstream of each shock reflection from the wall, the flow accelerated as indicated by the decrease in the static pressure. An additional static pressure decrease due to the Prandtl-Meyer expansion at the entrance of the diffuser throat is also apparent. Farther downstream the flow was decelerated from supersonic to subsonic flow by a shock structure sometimes referred to as a pseudoshock. This is indicated by the steady rise in the wall static pressure. The subsonic deceleration continued to the end of the divergent exit cone. For a stagnation pressure of 100 psia it appears that the length of the diffuser could be shortened to about six diffuser throat diameters without altering its performance. The shortened length would include the required convergent section at the entrance of the diffuser (Ref. 1) and the continuous pressure rise region farther downstream.

4. Summary and Conclusions

Wall static pressure measurements made in the upstream part of an axisymmetric second-throat diffuser

when started indicated a series of rather abrupt flow decelerations and gradual accelerations associated with an oblique shock at the inlet and its subsequent reflections from the wall and the centerline. Farther downstream the flow was gradually decelerated from supersonic to subsonic velocities as deduced from a continuous rise in the wall pressure. At a stagnation pressure of 100 psia it appeared that the minimum length required for the operation of this diffuser without loss in performance is about six throat diameters.

References

1. Massier, P. F., and Roschke, E. J., "Experimental Investigation of Exhaust Diffusers for Rocket Engines," *Progress in Astronautics and Rocketry*, Vol. 2, Liquid Rockets and Propellants, Academic Press, New York, pp. 3-75, 1960.
2. Massier, P. F., and Roschke, E. J., *Application of Exhaust Diffusers for Rocket Engine Testing*, Technical Memorandum 33-97, Jet Propulsion Laboratory, Pasadena, Calif., Aug. 21, 1962.
3. Baker, P. J., and Martin, B. W., "Some Operating Characteristics of the Supersonic Axi-Symmetric Parallel Diffuser," *J. Mech. Engr. Sci.*, Vol. 7, No. 1, pp. 15-22, 1965.
4. Back, L. H., Massier, P. F., and Gier, H. L., "Comparison of Experimental with Predicted Wall Static Pressure Distributions in Conical Supersonic Nozzles," *AIAA J.*, Vol. 3, No. 9, pp. 1606-1614, 1965.
5. Baker, P. J., "Heat Transfer in a Supersonic Parallel Diffuser," *J. Mech. Engr. Sci.*, Vol. 7, No. 1, pp. 1-7, 1965.

B. Suitability of a Hollow Cathode for a 20-cm-diam Ion Engine, E. V. Pawlik and D. J. Fitzgerald

1. Introduction

A clustered ion engine system is currently under study at JPL (Refs. 1 and 2) for use as primary spacecraft propulsion. At present, an oxide-coated cathode is utilized in the study. A hollow cathode, which requires much less cathode heating power and has greater lifetime capabilities, has been successfully incorporated in a 15-cm-diam flight type electron-bombardment ion engine (Ref. 3). It is of interest to determine if this type of cathode can be incorporated into the 20-cm electron-bombardment ion engines being used in the JPL electric propulsion system.

The tests described herein to evaluate hollow cathode operation were conducted in a bell jar. Tests were made to determine if this type of cathode could provide the 10-A emission current necessary for the larger thruster, to determine the cathode mercury flow necessary to provide this emission, and to evaluate the cathode lifetime capability. Total mercury flow to the thruster should be

about 9.6 g/h. The cathode should be able to deliver the required current at a small percentage of the total flow to the thruster in order to allow flexibility in both propellant injection method and output power level adjustment (for power matching to the solar cell output).

2. Apparatus and Procedure

a. Hollow cathode. Typical hollow-cathode construction is presented in Fig. 3. Cathode construction utilized a thin wall tantalum tube to which a refractory metal disc was electron-beam welded. A small orifice was located in the center of this refractory disk. A swaged heater was wound around the outside of the tantalum tube and a strip of tantalum foil, coated with a barium carbonate solution, was inserted in the tube. The heater, along with the low work-function surface which was beneficial in initiating a discharge, could readily provide electron emission at relatively low temperatures (1000°C). A keeper anode was located near the orifice to provide the electric field necessary to start the discharge and also to draw sufficient current to maintain it once established. Two types of hollow cathodes were used in this investigation. The cathode types differed mainly in the disk material at the orifice location. Tantalum and 2% thoriated tungsten were the materials used. Cathodes of this type have been used for both primary electron emission within the thruster and for ion beam neutralization (Refs. 3 and 4).

b. Test setup. A photograph of the bell jar experimental setup is shown in Fig. 4. The hollow cathode was mounted on a stainless steel frame provided with movable mounts for a keeper anode and for a second anode which was used to simulate the plasma within a thruster. The mer-

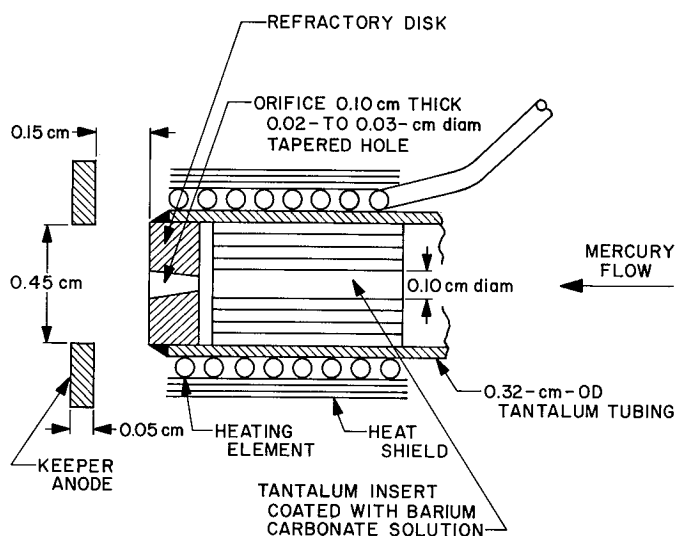


Fig. 3. Typical hollow-cathode construction

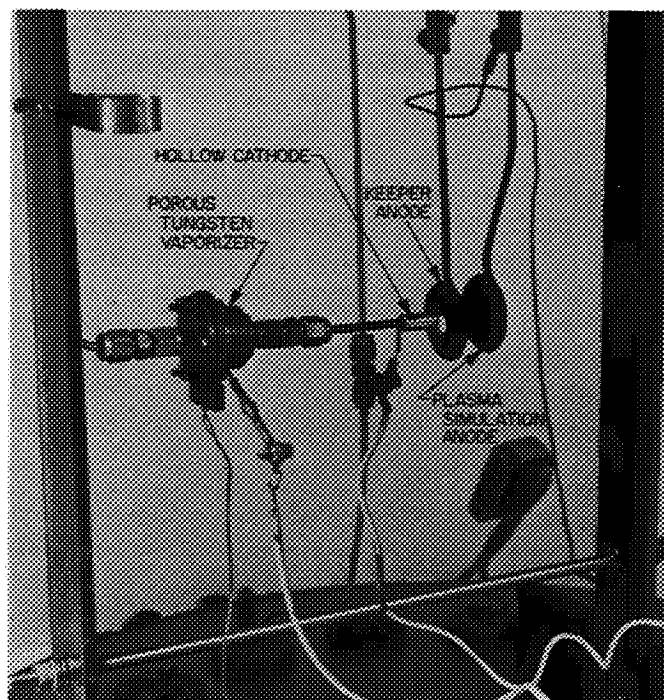


Fig. 4. Experimental hollow-cathode setup within a bell jar

cury was fed from a 1-mm-diam pipet, which was located outside of the bell jar. Mercury flow was determined from the measured rate of mercury level change during cathode operation.

An electrical schematic diagram for the test setup is shown in Fig. 5. Current limiting resistors were placed in

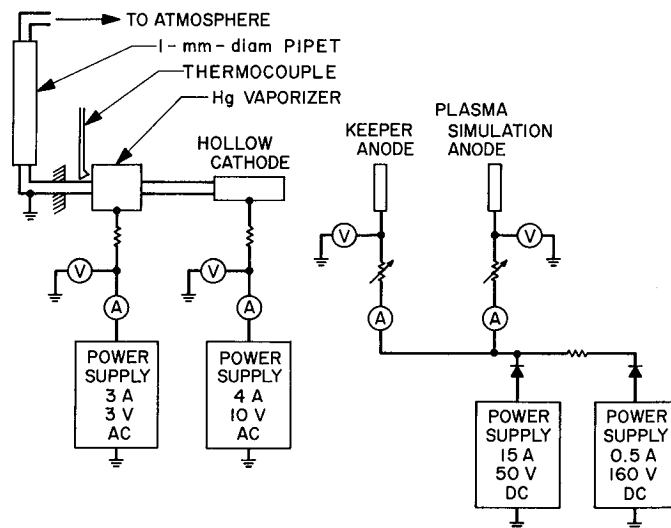


Fig. 5. Schematic of hollow-cathode bell jar test setup

both anode electrical circuits to reduce the starting voltage as the discharge current increased and to provide control during cathode operation.

Cathode characteristics were obtained at various values of emission current and cathode-to-plasma-anode distances by recording the anode voltages over a range of flow rates. Microphotographs of the orifice were taken before and after each test to document orifice erosion.

3. Results and Discussion

In general, two problems of operation with a hollow cathode at emission current levels of 5 A, or greater, were much more severe than at lower levels of emission. These were: (1) sputtering damage at voltages of 30 V or greater was more intense, and (2) cathode temperatures were very high. The high temperatures were probably due to ion bombardment and were found to be on the order of 2000°C. Sputtering damage was found to rapidly

close the orifice when the voltage was above 30 V. Therefore, it was found necessary to minimize operating time in the region between 30 and 40 V and to avoid entirely higher voltages during operation.

Initial cathode tests were made using tantalum for the refractory disk in which the orifice was located. These cathodes usually started and operated at lower mercury flows than the tungsten cathodes (Fig. 6). The tantalum cathode, however, was more susceptible to sputtering damage and melting, while restart capabilities were noted to be erratic, and in several cases the discharge could not be reestablished. Because of these difficulties it was decided to concentrate on the more promising thoriated-tungsten type.

The tungsten hollow-cathode shows some promise of being able to operate for sustained periods at currents on the order of 10 A. Data has been obtained for up to 15 A

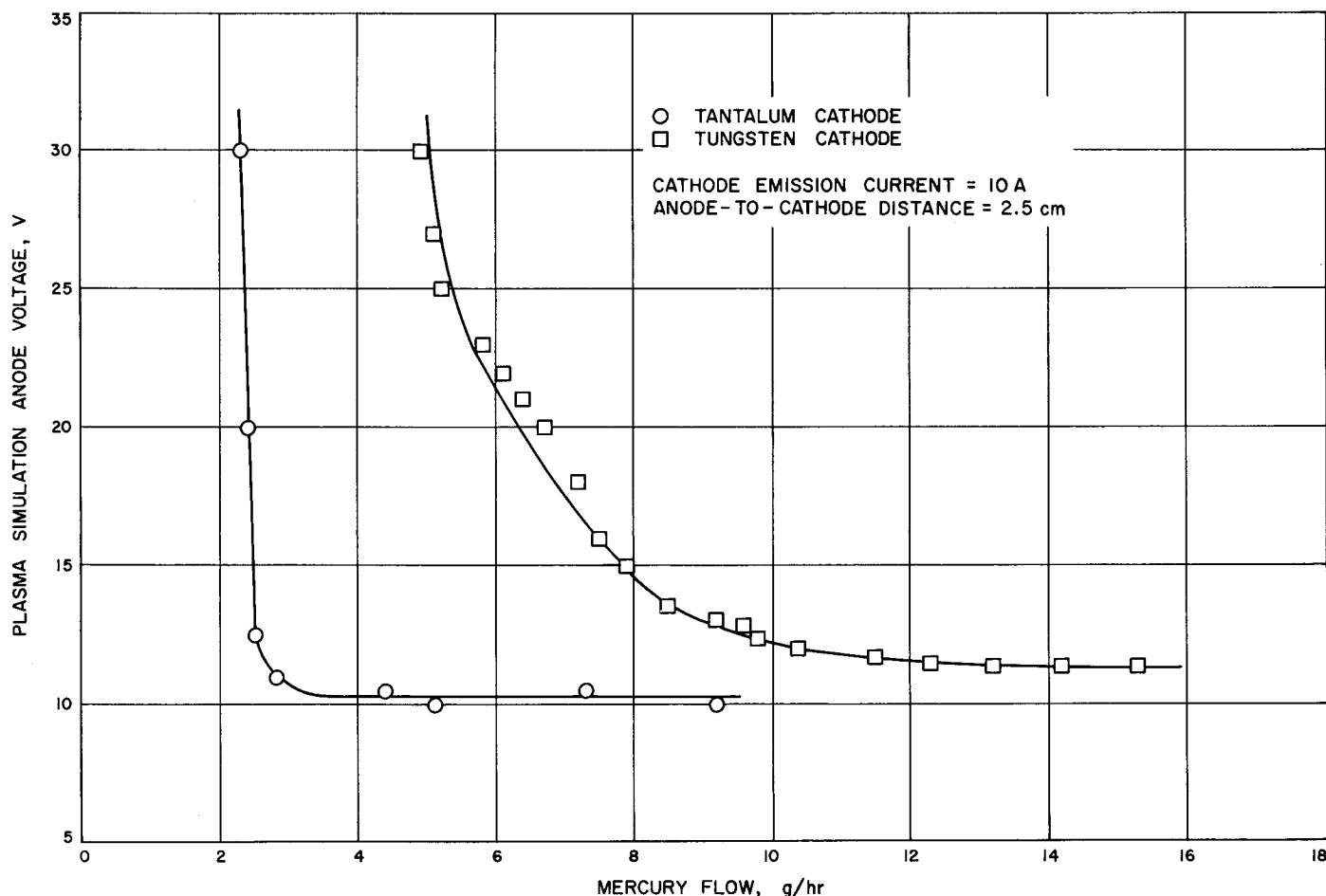


Fig. 6. Comparison of tantalum and tungsten cathode operation

of emission current. Microphotographs indicate some melting of the material near the downstream edge of the orifice. Photographs are presented in Fig. 7 for 6 h of cathode operation at 10 A of emission current. Slighter amounts of melting were also noticed after sustained periods at 5 A of emission current. Progressive interruption and inspection of the cathode during tests indicate that the major portion of the deterioration seemed to occur within the first few hours of operation. The orifice apparently deteriorates until an optimum size and shape are achieved. The final size arrived at, after a few hours,

was apparently in thermal balance with the plasma heating. Unfortunately, there were neither enough time nor facilities to make a long-term lifetime test of the cathode; therefore, the discussion is somewhat speculative. A larger orifice size, closer to the observed optimum size, appears to be desirable. The lifetime of a larger orifice size would have to be determined. For long lifetimes to be obtained, the cathode should remain unchanged in appearance for any operating time that represents a small fraction of the desired lifetime.

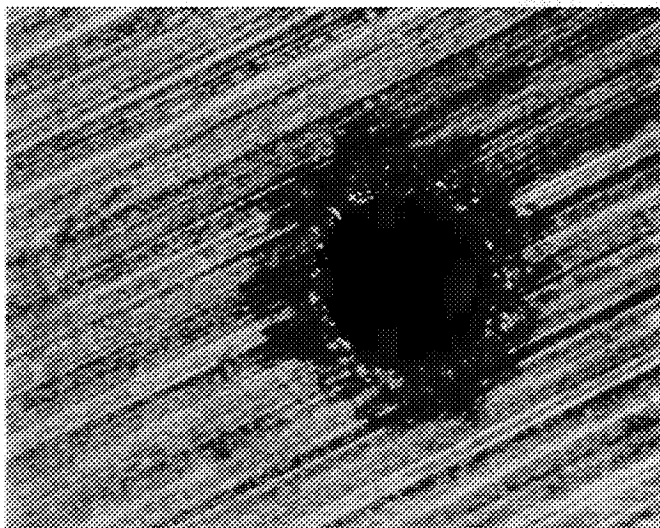
An alternative approach, suggested by the slighter amounts of melting at lower emission currents, would be the use of multiple cathodes. This approach might provide cathode lifetime at a cost in system complexity.

The characteristics of the hollow cathode were obtained by recording the cathode-to-plasma-anode voltage as the mercury flow was varied. These characteristics were obtained at several values of anode spacing and emission current. The keeper current and voltage were not varied during these tests. Keeper voltages on the order of 10 V were maintained.

Figure 8 is a plot of the cathode mercury flow versus cathode-plasma voltage for three values of emission current. The mercury flow necessary to maintain the 5-A emission was noted to require only a slightly lower flow than that necessary to maintain 10 A. The flow necessary to maintain 15 A was observed to be approximately equal to that necessary to maintain 10 A. This data was obtained in what has been referred to as the spot mode of operation (Ref. 4). If the emission current is reduced below about 3 A, a change of modes was usually observed.

It has been found desirable to maintain the cathode-plasma voltage at a relatively high value (Ref. 3), since the thruster performance improves as this voltage approaches 40 V. A practical limit exists on this voltage inasmuch as the cathode sputtering increases considerably at voltages higher than 30 V. If the cathode-plasma voltage is held at 30 V, the mercury flow to maintain the desired 10-A emission is on the order of 5 g/h which represents 52% of the 9.6 g/h total mercury flow to the thruster. Small increases in this flow rate could cause large increases in the emission current. Since proper operation of the thruster at different power levels requires the voltage be held relatively constant while changing the emission current, there could be a problem in matching the desired emission at a set of flow rates if the cathode operates on a fixed percentage of the total mercury flow

(a) BEFORE OPERATING



(b) AFTER 6 h OF OPERATION AT 10-A EMISSION CURRENT

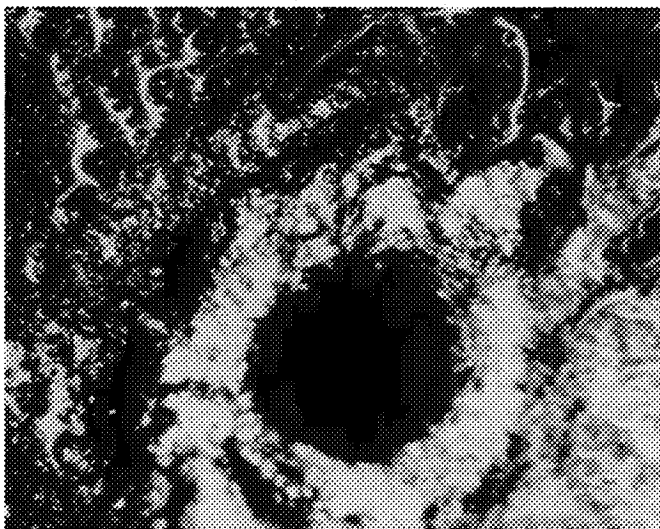


Fig. 7. Microphotograph of downstream side of hollow-cathode orifice (0.033-cm diam)

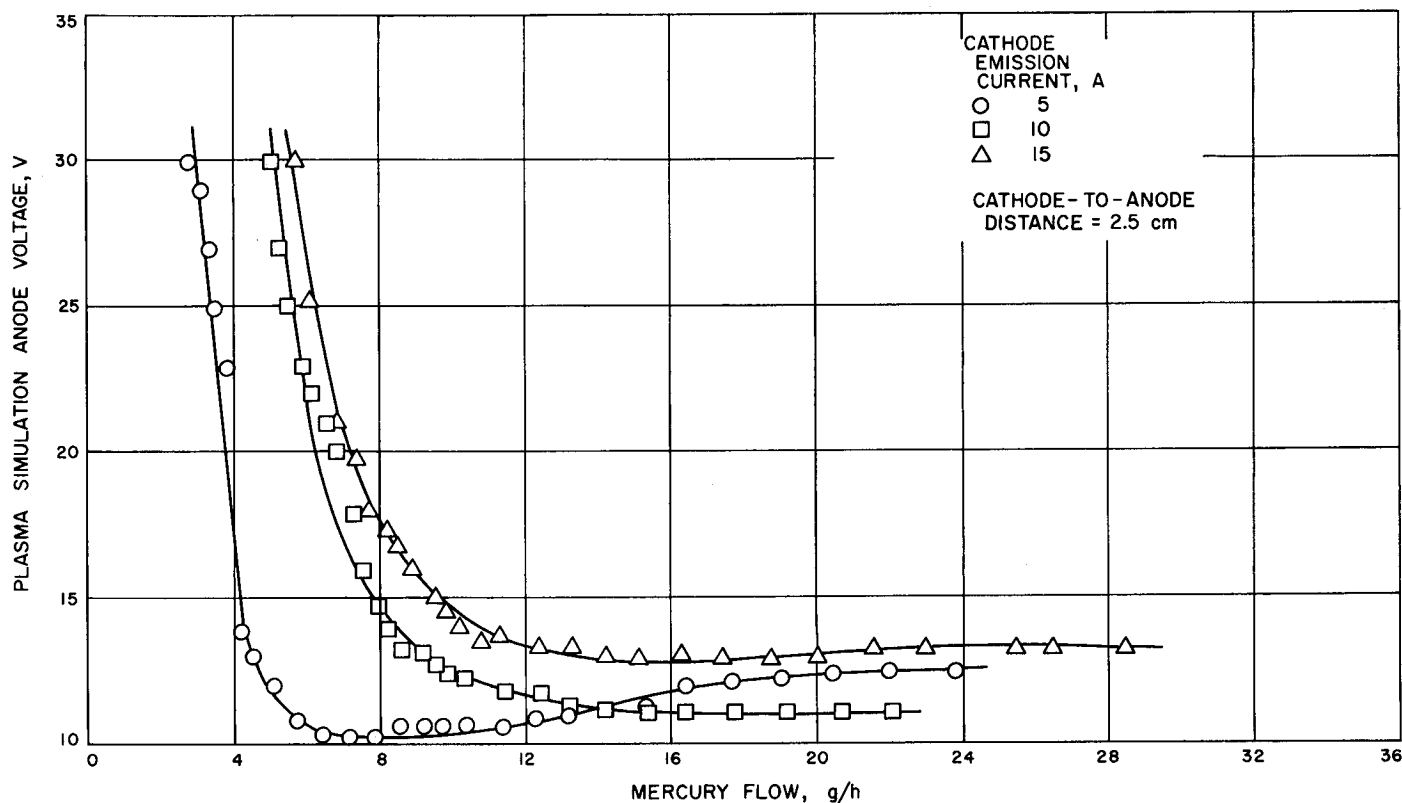


Fig. 8. Effect of cathode emission on operation of tungsten hollow cathode

to the thruster. A separate flow to the cathode in addition to the flow to the thruster may be possible, but complex system logic or thruster inefficiencies may result.

Figures 9 and 10 present two sets of curves, each at constant emission current for various cathode-to-plasma-anode distances. Both sets show quite clearly that the lowest flow operation is obtained when the anode is close to the cathode. This shows that the flow rate required for proper cathode functioning is very sensitive to anode location. The cathode flow will be reduced if the plasma within the thruster is located close to the cathode. The cathode heating might also be considerably reduced if the potential distribution in this region is altered significantly. Any further evaluation of this cathode type should be conducted either in an operating thruster or in a thruster ion chamber without extraction grids.

4. Conclusions

It was determined that a hollow cathode could deliver emission currents of up to 15 A at cathode mercury flows of 5 g/h or less, depending on the exact plasma boundary location. Thoriated tungsten was found to be the more

suitable of the two materials tested, but some melting of the surface around the orifice was noted.

The bell jar tests indicate that a slightly larger orifice should be considered, and also that testing in an operating thruster, or in a thruster ion chamber, is necessary in order to more fully evaluate cathode lifetime. Power matching of the thruster by means of mercury flow modulation could be problematical with this type of cathode.

References

1. Kerrisk, D. J., and Kaufman, H. R., "Electric Propulsion Systems for Primary Spacecraft Propulsion," AIAA Paper 67-424, presented at AIAA Conference, Washington, D.C., July 1967.
2. Masek, T. D., and Womack, J. R., "Experimental Studies with a Clustered Ion Engine System," AIAA Paper 67-698, presented at AIAA Conference, Colorado Springs, Colo., Sept., 1967.
3. Kerslake, W. R., Byers, D. C., and Staggs, J. F., "SERT II Experimental Thruster System," AIAA Paper 67-700, presented at AIAA Conference, Colorado Springs, Colo., Sept., 1967.
4. Rawlin, V. K., and Pawlik, E. V., "A Mercury Plasma-Bridge Neutralizer," AIAA Paper 67-640, presented at AIAA Conference, Colorado Springs, Colo., Sept., 1967.

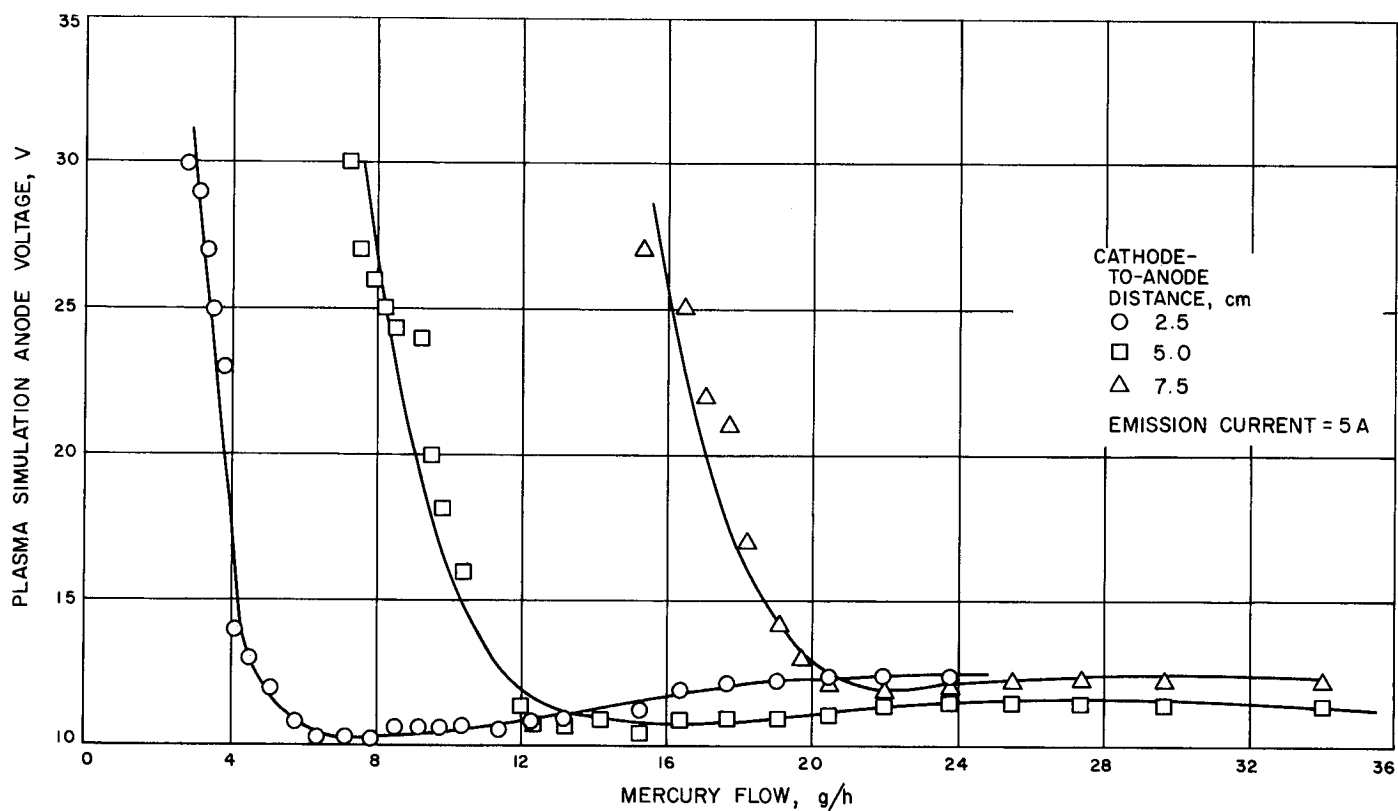


Fig. 9. Effect of cathode-to-anode distance on tungsten hollow cathode

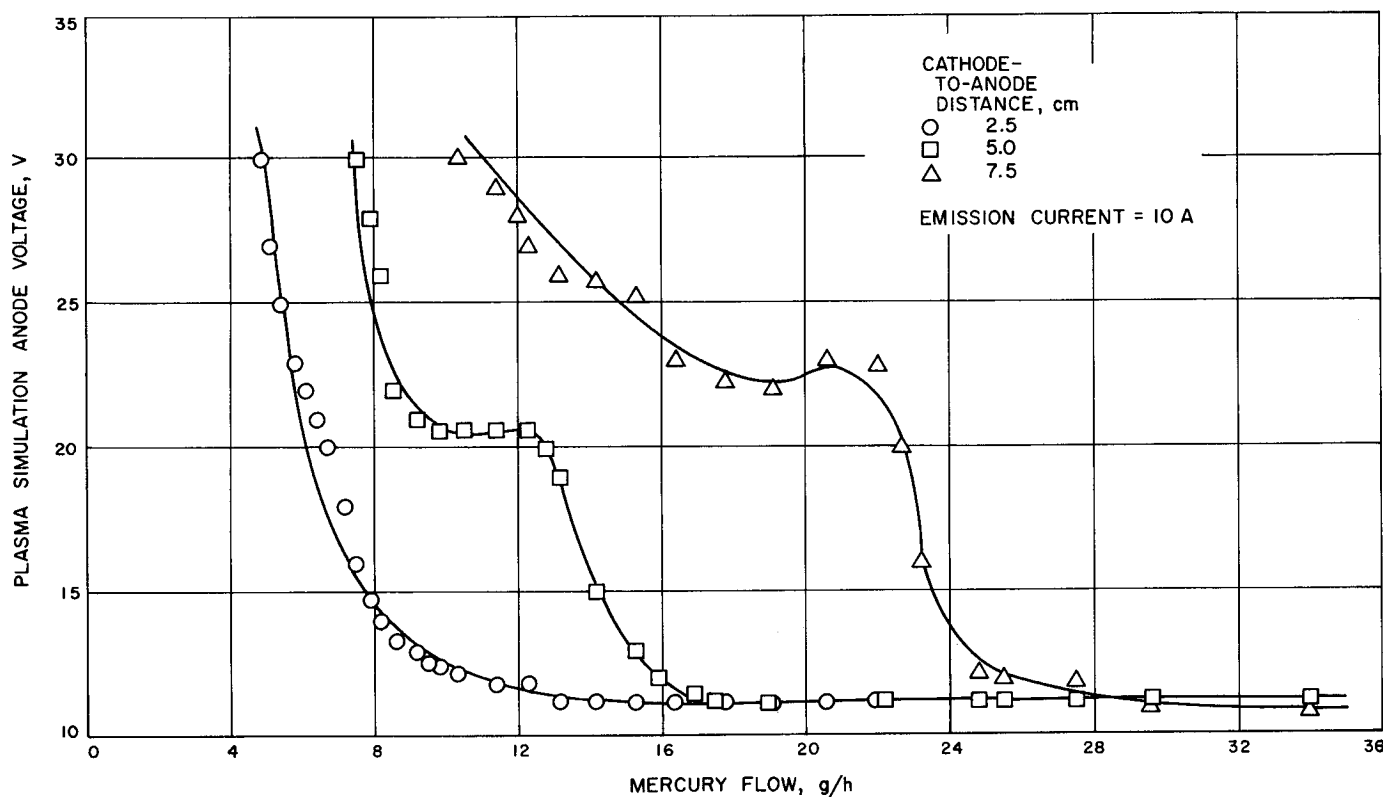


Fig. 10. Effect of cathode-to-anode distance on a tungsten hollow cathode

C. Liquid-Metal MHD Power Conversion,

D. G. Elliott, L. G. Hays, and D. J. Cerini

1. Introduction

Nuclear-electric powerplants for electric propulsion will be required to have operating lifetimes of at least 7000 h. Nonrotating power conversion systems may achieve this lifetime most easily. A nonrotating Rankine cycle being investigated is the liquid-metal magnetohydrodynamic (MHD) system wherein lithium is accelerated by cesium vapor in a two-phase nozzle, separated from the cesium, decelerated in a magnetohydrodynamic generator, and returned through a diffuser and heat source to the nozzle.

The experiments currently being prepared are a cesium-lithium loop for erosion and small scale component testing at 2000°F and a 50-kW conversion system for testing with NaK (in place of lithium) and nitrogen gas (in place of cesium vapor) at room temperature. Design of a cesium-lithium conversion system and construction of the test facility have begun.

2. High-Temperature Experiments

a. Cesium-lithium erosion loop. Fabrication of the 100 kWt erosion loop has continued. The niobium-1% zirconium flow system will be operated at 2000°F (1100°C) at lithium velocities of 500 to 650 ft/s (150 to 200 m/s). The thermodynamic cycle will be the same as that of a liquid-metal MHD space power system without a generator. A schematic of the flow system and values of the design operating conditions are shown in Fig. 11. Initial experiments to be performed will include erosion and corrosion of separator and flow channel materials at lithium impact and flow velocities of 500 to 650 ft/s, performance of a two-phase nozzle with cesium and lithium, and condensing heat-transfer coefficients for cesium vapor with lithium added.

The two-phase nozzle, cesium condenser, and lithium heater have been completed. The nozzle was tested with nitrogen and water to verify agreement with the nozzle computer program for an immiscible fluid pair with accurately known properties. The nozzle is shown during

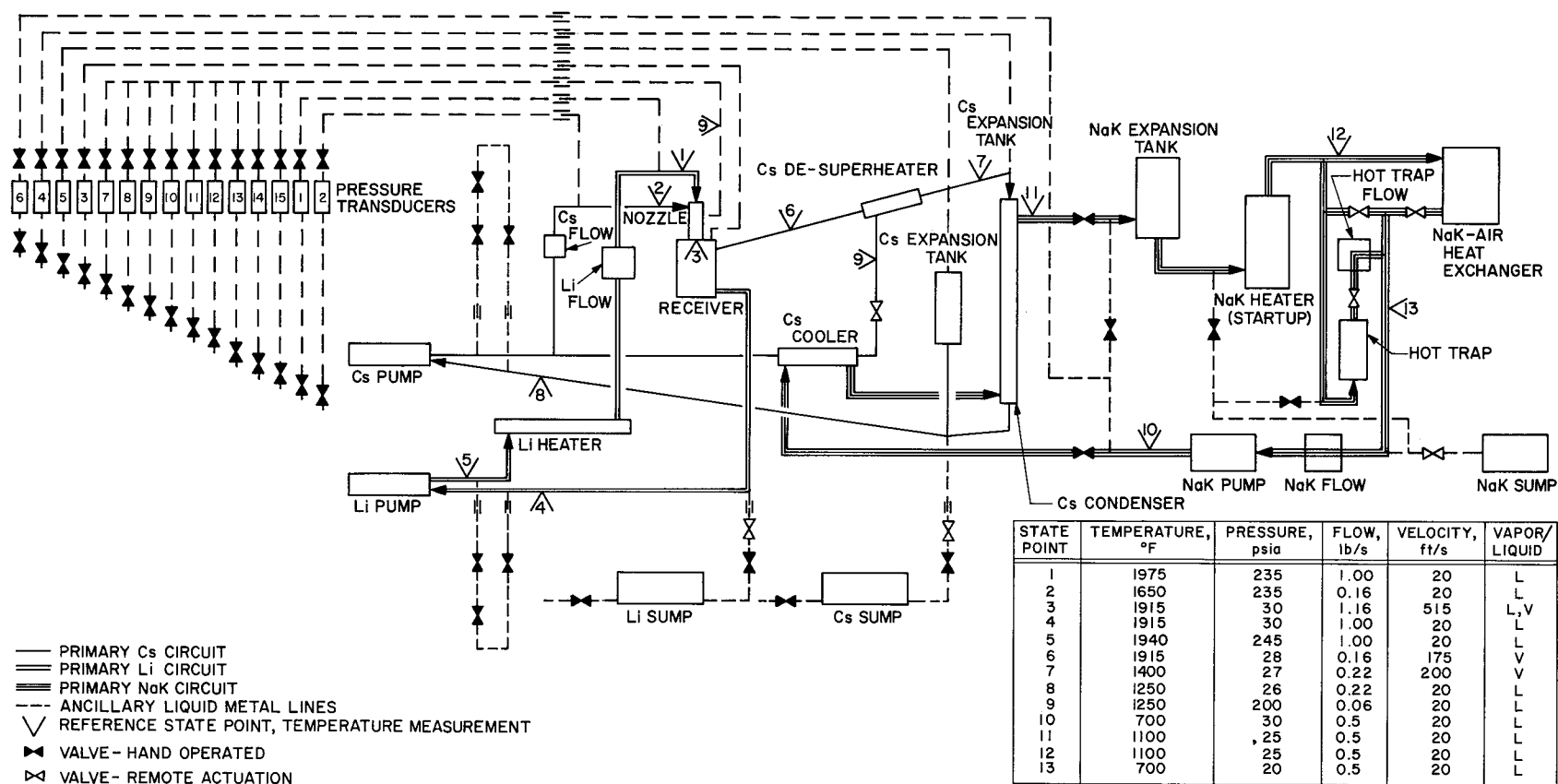


Fig. 11. Cesium-lithium erosion and component test loop

test in Fig. 12. The exit velocity determined by thrust measurement agreed within 5% with the calculated velocity as shown in Fig. 13. The mass ratio of lithium to cesium at the loop operating point is 6.25, and the calculated exit velocity is 600 ft/s.

b. Cesium-lithium conversion system. Design of a 200 to 300 kWe liquid-metal MHD power system and test facility to operate at 1800°F (980°C) is continuing. The facility piping and equipment arrangement selected is shown in Fig. 14. The locations of the 50 kWe NaK-nitrogen conversion system and 100 kWt erosion experiment are also shown.

3. NaK-Nitrogen Conversion System

a. Fabrication status. The modifications following the water-nitrogen calibration tests have been completed.

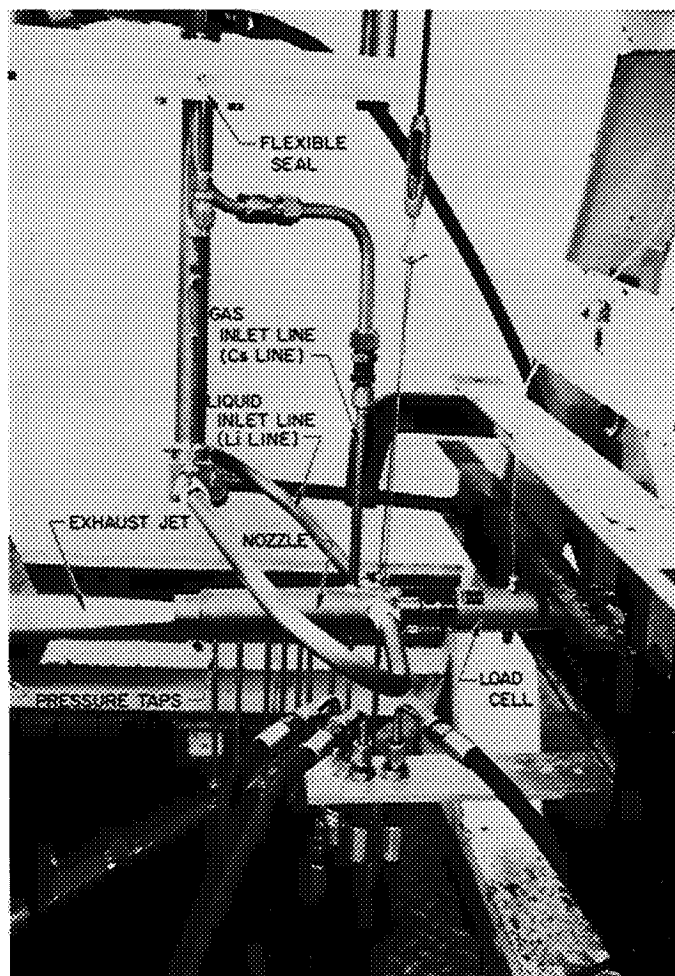


Fig. 12. Water-nitrogen test of nozzle for cesium-lithium loop

The modifications are: 5-in. separator extension, separator side-wall flow diverters, elimination of side-wall exhaust ports, and increase of nitrogen exhaust flow area to 15 in.² The separator is now being assembled with the first set of generator stator blocks. Installation of the NaK tanks, NaK piping, nitrogen bottle bank, main nitrogen line, and nitrogen pressurizing system have been completed.

b. Hydraulic test evaluations. The flow velocity at the separator exit (upstream diffuser inlet) was summarized in SPS 37-47, Vol. III, p. 132, Fig. 34. The exit velocity was calculated as the ratio of the velocity thrust to the sum of the gas and liquid flow rates, where the velocity thrust was the measured thrust reduced by the pressure thrust. The pressure thrust was the product of the exit pressure and exit area of the separator, using the average of four pressure readings across the capture slot. A typical pressure distribution was 1.5 atm across the upper half of the slot, approximately the same as the nozzle exit pressure, and a linear rise from the flow centerline to 10 atm at the separator surface. This pressure can be explained by assuming that the liquid turns more sharply than the actual separator surface at the exit. A 3-in. radius of curvature would produce a centrifugal force corresponding to the observed surface pressures. If, in addition,

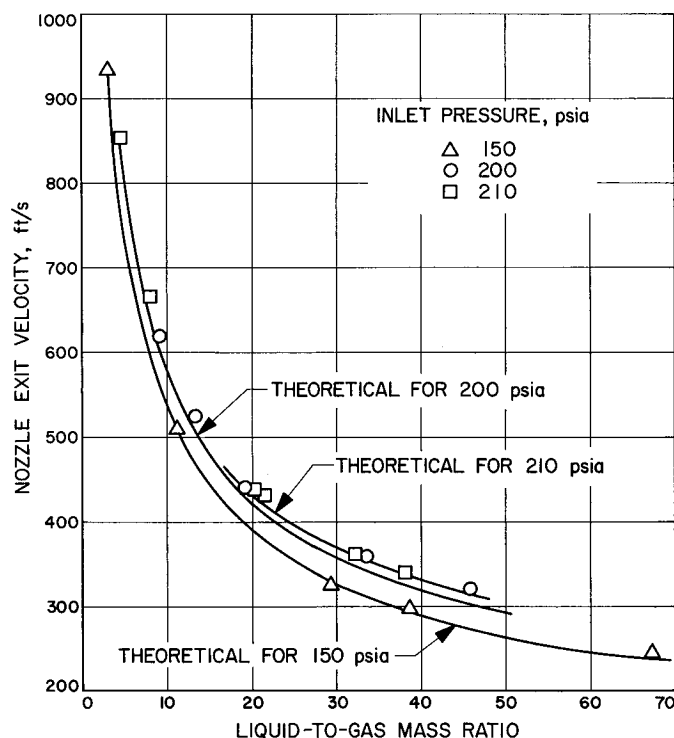


Fig. 13. Comparison of experimental and theoretical exit velocities for cesium-lithium loop nozzle operating with nitrogen and water

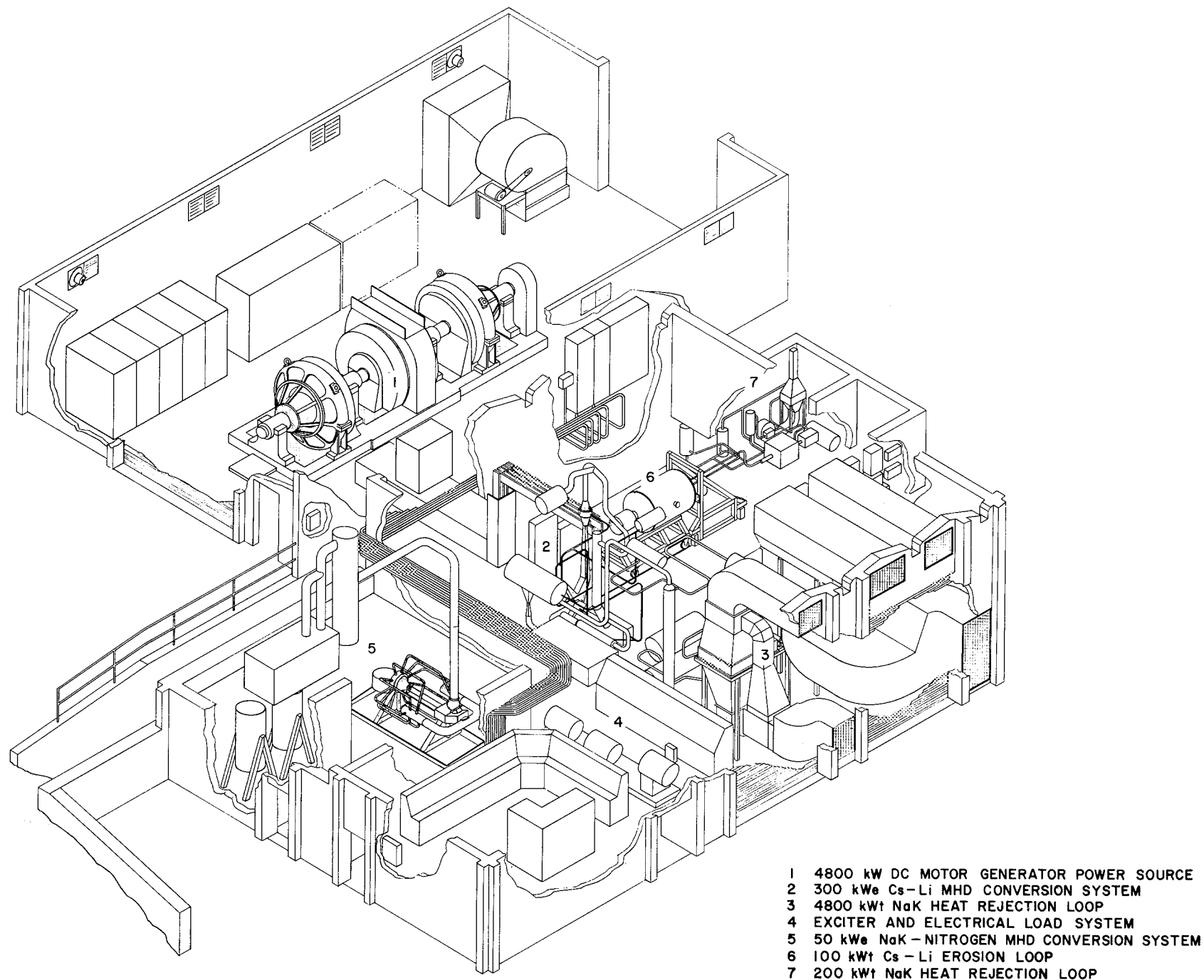


Fig. 14. Facility arrangement for 1800°F MHD conversion system experiment

the liquid film is assumed to have 0.25 void fraction of gas, agreement is obtained with the observed pressure gradient away from the surface. Using the mean liquid pressure from this model, the separator exit velocities were corrected to the higher values corresponding to no pressure rise and inserted in the separator loss relations (SPS 37-27, Vol. IV, pp. 76 and 77) to calculate a value of measured skin friction coefficient. The ratio of these values to the high-Reynolds number flat-plate skin friction coefficient, $C_f = 0.026 \times (\text{Reynolds number based on mean film thickness})^{-0.2}$, was an average of 1.2, showing the liquid film forms an essentially fully developed turbulent boundary layer.

The velocity and pressure at the separator exit (upstream diffuser inlet) were used with the velocity and

pressure at the upstream diffuser exit to calculate the diffuser efficiency, which is the ratio of the observed to the isentropic pressure rise. The isentropic pressure rise was evaluated assuming the liquid film of 0.25 void fraction first accelerates to the mean separator exit pressure, then undergoes an isentropic compression to the measured diffuser exit velocity. The results are shown in Fig. 15. Within the 3% uncertainty in diffuser exit velocity, the diffuser can be described by a single curve of efficiency versus inlet volume ratio, and satisfactory efficiency is obtained at the volume ratios below 1.0 to be employed in the conversion system tests.

D. Efficiency of Thermionic Diodes at Reduced Power Output, J. P. Davis

1. Introduction

A typical power profile for unmanned planetary nuclear electric propulsion missions involves an initial full-power thrusting period, a nonthrusting coast mode, and a second full-power thrusting period. The coast mode is characteristically about 40% of the total flight time. During this coast mode, power is required for several spacecraft functions plus self-sustaining power for the propulsion plant. The zero thrusting power requirement, or hotel load, is usually estimated in the neighborhood of 5 to 10% full power. It is desirable that the nuclear plant be capable of supplying this power at some reasonable fraction of its full power design efficiency in order to minimize fuel burnup during the coast mode. Some concern has been exhibited for the performance of a thermionic reactor at power levels considerably reduced from its design point, particularly with respect to efficiency.

From the viewpoint of spacecraft and plant power demands, it would appear highly desirable to operate the thermionic reactor in a constant voltage output mode at all power levels. In this fashion variable-turns-ratio transformers, or their equivalent, can be avoided and constant voltage supplied to all on-line electrical limits.

2. Studies

Initial reactor control studies at JPL¹ have indicated that constant voltage control is indeed feasible and rather straightforward, both for perturbations about the nominal operating point and for large power demand changes. The question of efficiency, however, was not considered in

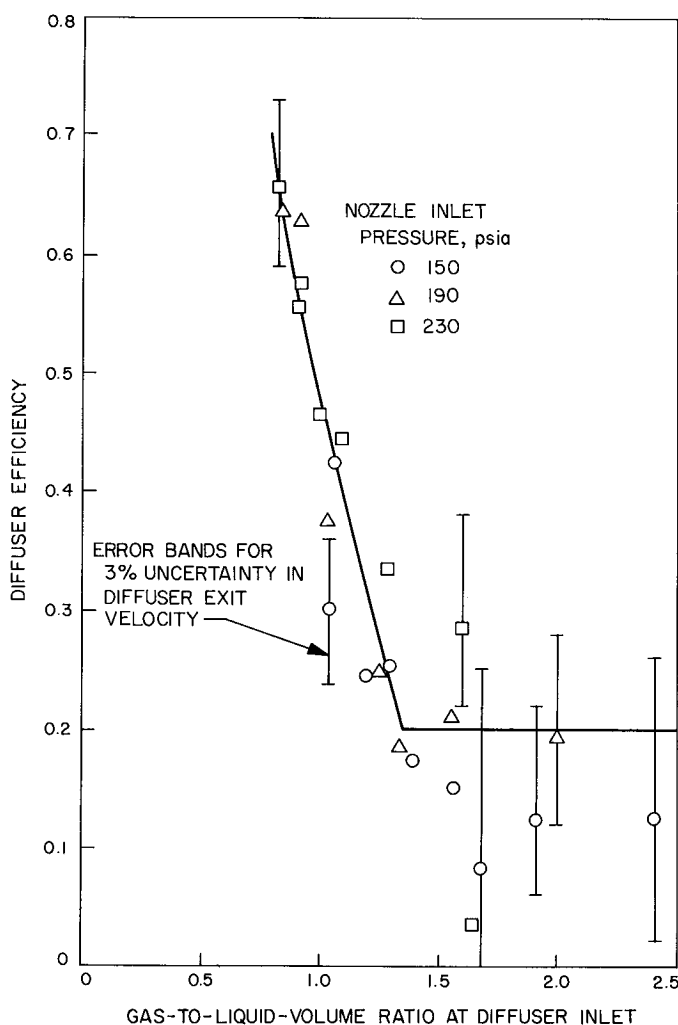


Fig. 15. Efficiency of the upstream diffuser for the NaK-nitrogen conversion system

¹Gronroos, H., Weaver, L., Guppy, J. G., and Davis, J. P., "Control System Design for an In-Core Thermionic Reactor" (to be published).

these control studies, and the purpose of this paper is to determine the efficiency penalty associated with constant voltage control as opposed to optimum matching of diode characteristics at reduced power. For purposes of this investigation, the following assumptions were made:

- (1) G.E. Simcon basic J - V data (extrapolated to temperatures below 1700°K), W emitter, Mo collector, 0.010-in. gap, optimum Cs temperature.
- (2) Fixed collector temperature at 1000°K .

A diode design was postulated to permit calculation of emitter-collector axial voltage drop. All other parameters are essentially independent of design details. Preliminary survey of output power and efficiency at the operating temperature of 2000°K with lead optimized for each output point permitted selection of the nominal full power operating conditions. These values and other key parameters are tabulated below:

Nominal Operating Point

14 A/cm², 0.667 V, 9.35 W/cm², 14.2% efficiency.
Unconditioned net power after emitter-collector and lead losses were subtracted.

Maximum Efficiency Point

15.3% at 10 A/cm² ($P = 8.14 \text{ W/cm}^2$).

Maximum Power Density Point

9.75 W/cm² at 18 A/cm² ($\eta = 12.9\%$).

Nominal Operating Point Losses

Axial emitter-collector voltage drop = 0.070 V.
Optimized lead voltage drop = 0.08 V.
Optimized lead thermal conduction loss = 7.8 W/cm^2 .
Thermal radiation loss = 11.6 W/cm^2 ($\epsilon_e = \epsilon_c = 0.3$).
Thermal cesium conduction loss = 2.5 W/cm^2 .
Thermal electron cooling loss = 43.0 W/cm^2 .

Power Desired During Coast

10% nominal full power

The nominal operating point was selected as a compromise between peak efficiency and peak power density. (A criticality limited core might be designed closer to peak power density, and a burnup limited core might be designed closer to peak efficiency.)

All losses are recomputed at various J - V locations and emitter temperatures to construct the output current density versus net output current curves shown in Fig. 16.

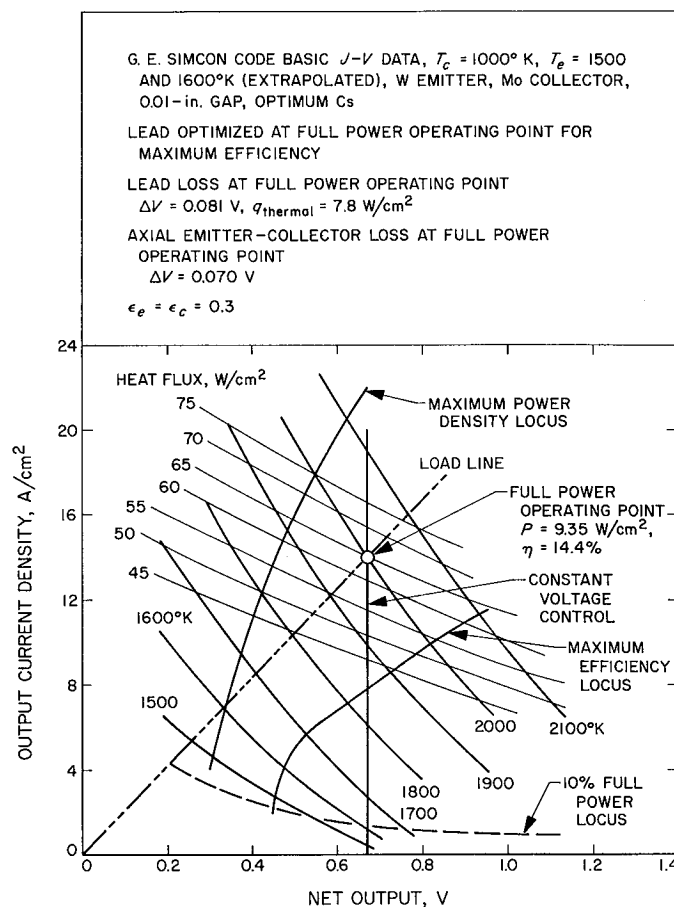


Fig. 16. Output current density versus net output voltage at various emitter temperatures

The locus of the 10% full power requirement (0.935 W/cm^2) is indicated. The loci of maximum power density and maximum efficiency are also shown. Figure 17 shows the efficiency curves replotted against net output voltage with the loci of maximum efficiency, 10% full power, and constant voltage output superimposed. In order to supply 10% full power at a voltage equal to the full power voltage, emitter temperature falls to $\sim 1600^{\circ}\text{K}$, and efficiency falls to 6.7%. If the constraint of constant voltage is removed and efficiency maximized, the result is $\sim 1500^{\circ}\text{K}$ emitter temperature at 7.4% efficiency.

The surprising result is the relatively minor efficiency penalty paid for constant voltage control as opposed to that attainable at the optimum voltage output of 0.45 V. The emitter temperature is already 400°C lower than the nominal operating temperature; so little longevity incentive exists for further temperature reduction of 100°C or so. Also, the magnitude of the efficiency, 6.7%, is encouragingly high, relative to the full power operating efficiency

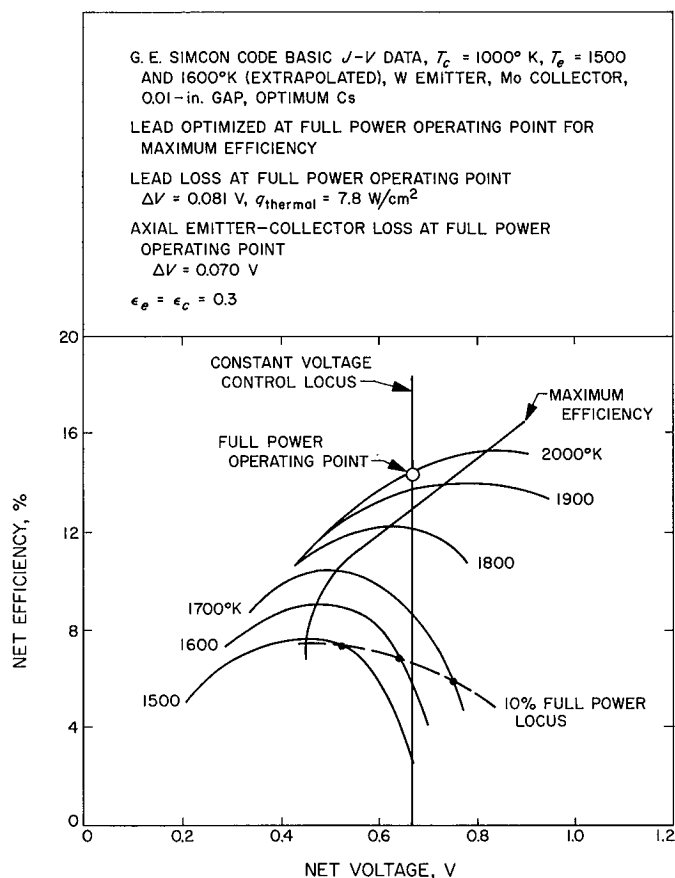


Fig. 17. Net efficiency versus net output voltage at various emitter temperatures

of 14.4%. A constant speed turboalternator efficiency would fall at least this much at one tenth the design power. It therefore appears that only minimal efficiency penalties are incurred by constant voltage control over optimal voltage control down to as low as 10% full power.

Full power operation at the peak power density point results in a somewhat more favorable reduced power operating point under constant voltage control, but it appears that the improvement ($\eta = 7.3\%$ versus $\eta = 6.7\%$) is not significant enough to warrant the loss in full power flexibility with respect to diode failures and/or performance degradation resulting from operation at the peak power density point.

Three cautionary statements are in order:

- (1) Fixed collector temperature was assumed, implying radiator area and/or radiator flow variation. Either or both of these modes are possible but involve additional system considerations. Reduction in coolant, and hence collector, temperature should also

be investigated as probably the simplest system operating mode. Diode efficiency, however, would suffer from reduced collector temperature.

- (2) The 1500 and 1600°K emitter temperature curves were extrapolated, and substantial uncertainties undoubtedly exist. The effect of these uncertainties on the results, however, is more in the direction of modifying the equilibrium operating temperature and less on the resulting efficiency.
- (3) At low emitter temperatures, peculiar inversions and multivalued J - V curves at fixed cesium conditions exist at low current densities. Diode output stability in this region may pose some further problems.

E. Clustered Ion Engine Systems Studies,

T. D. Masek

1. Introduction

Solar electric propulsion systems are presently being considered for future spacecraft primary propulsion (Refs. 1-3). The major elements and performance requirements for such systems were discussed in Ref. 1. Since a demonstration of the feasibility of using solar electric systems requires a test of realistic hardware, an experimental system development program has been initiated. Preliminary tests of a clustered thruster and propellant tankage system were reported in Ref. 4. This system is shown schematically in Fig. 18 and photographically in Fig. 19.

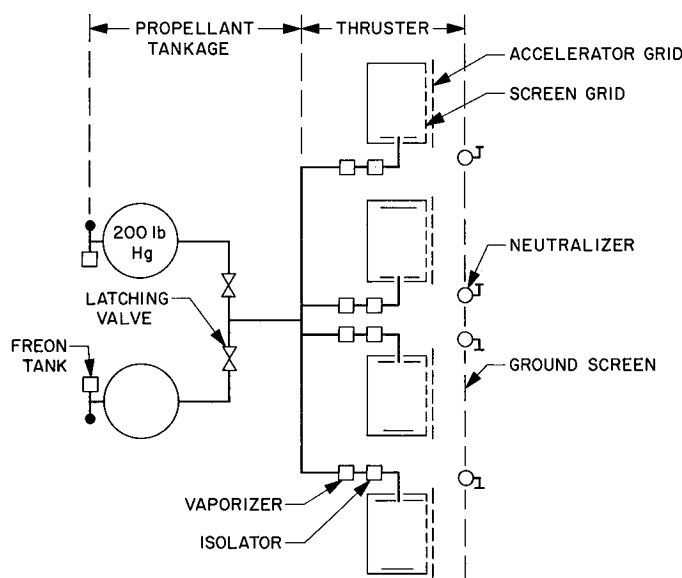


Fig. 18. Clustered ion thruster system schematic

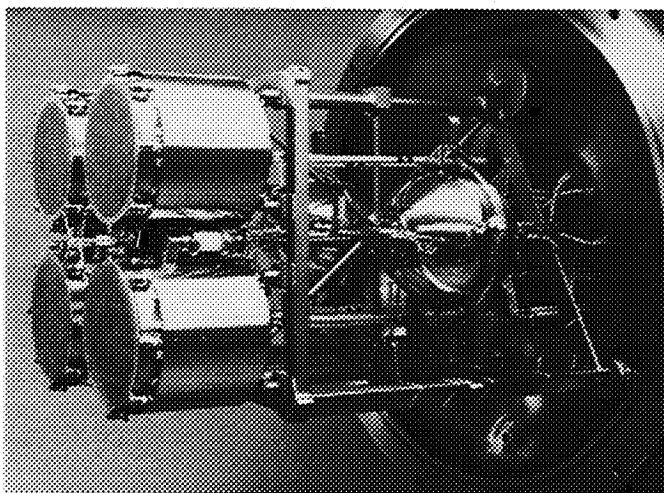


Fig. 19. Clustered ion thruster system

This report describes additional tests using the system of Ref. 4. Tests of the thruster subsystem were conducted to: (1) show methods of improving system performance, (2) determine limits of off-design-point thruster operation, and (3) determine power conditioning and control require-

ments. Propellant tankage subsystem tests were conducted to determine volume requirements for the Freon pressurization system and to study possible bladder deterioration due to temperature and pressure cycling.

2. Thruster Operation

The most significant parameter in discussing thruster performance is the ratio of discharge power to beam current. The values of this ratio for the 20-cm engines reported in Ref. 4 were 650 to 750 eV/ion. The 15-cm SERT II thruster design, using permanent magnets, was reported to operate in the range of 150 to 250 eV/ion (Ref. 5). Consideration of the SERT II design suggested a number of geometry modifications which might improve the 20-cm thruster. However, the effect of these modifications on a 20-cm thruster was not directly apparent from Ref. 5, and a detailed series of systematic changes in the existing design was required.

The changes producing the greatest performance improvements were: (1) the addition of an iron cylinder around the cathode (cathode pole piece), (2) movement

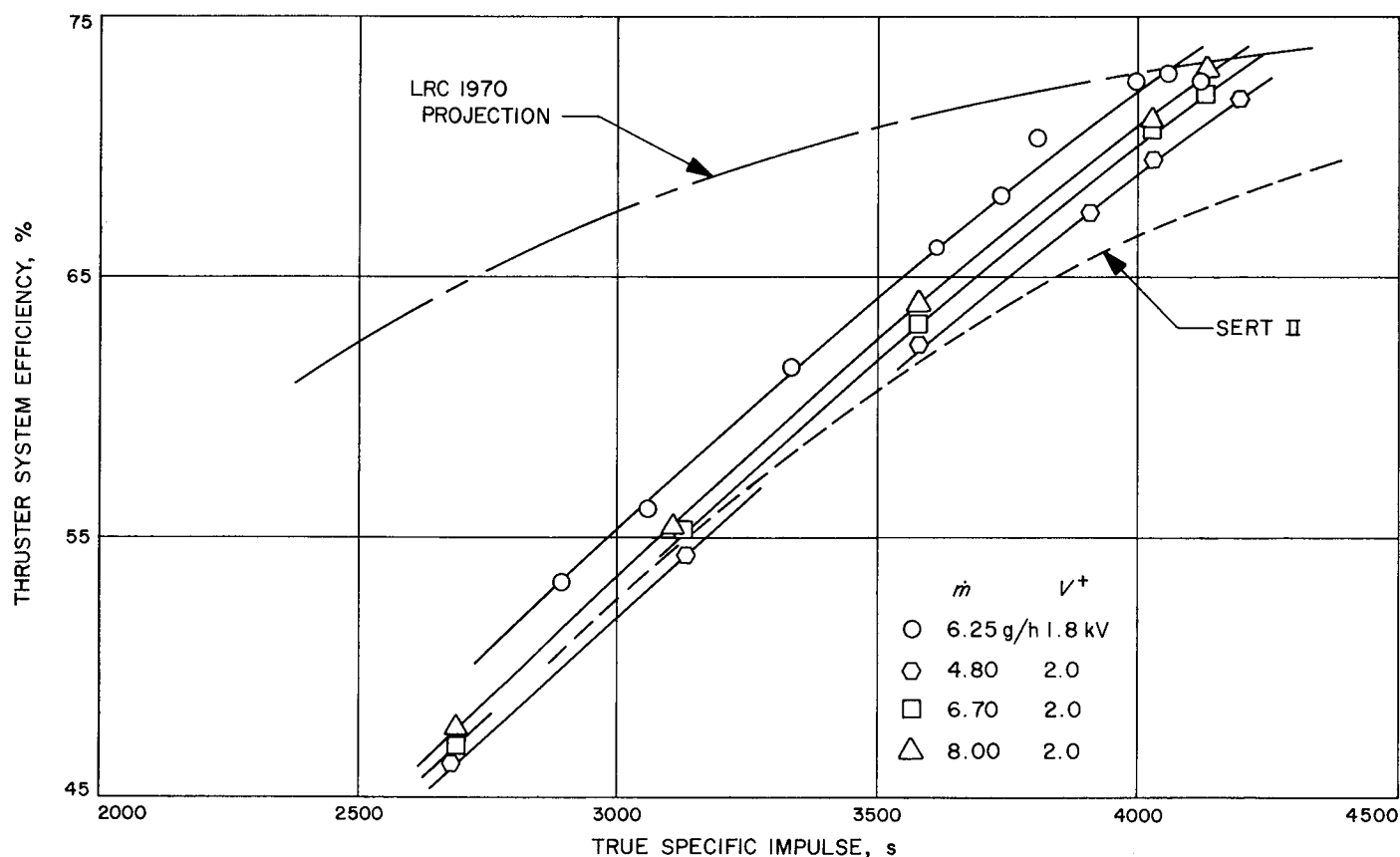


Fig. 20. Effect of flow rate on discharge power

of the cathode to the forward end of the pole piece, and (3) a reduction of the screen grid thickness from 0.100 to 0.052 in. These changes caused reductions in discharge power of approximately 125, 175, and 150 eV/ion, respectively. Thus, the discharge power has been reduced to 250 to 350 eV/ion.

A number of other factors, not previously reported, were also found to affect the discharge power losses. These are: (1) propellant flow rate, (2) discharge voltage, and (3) net ion acceleration voltage. The magnitudes of these factors at 80% mass utilization efficiency are, respectively (using + for increases in eV/ion with increases in the parameter): (1) + 15 eV/ion per g/h, (2) + 2.5 eV/ion/V (30 to 40 V discharge) and (3) - 12.0 eV/ion/100 V (nominal $V^+ = 2.0$ kV, $V^- = 2.0$ kV). In addition, heating the spiral "flower cathode" with ac current increases the discharge eV/ion by about 20, compared with optimum dc heating.

Discharge eV/ion for a thruster with the modifications described above is shown in Fig. 20 for several propellant flow rates. Shown for reference is a curve for the equivalent

SERT II flow rate, that is, the flow rate obtained by directly scaling from the 15-cm SERT II thruster design point to the 20-cm thruster used in this work.

The current striking the accelerator grid (interception current) relates directly to grid life, since this current results in sputtering away of grid material. A plot of the ratio of interception to beam current is shown in Fig. 21. The figure shows that grid life depends directly on propellant flow and mass utilization efficiency.

Thruster total efficiency data, including vaporizer power, but not including neutralizer power, is shown in Fig. 22. A curve for SERT II and a Lewis Research Center projection for 1970 are presented for reference (Ref. 1).

3. Propellant Tankage

Tests performed on the propellant tankage system since those reported in Ref. 4 produced two basic conclusions. First, the liquid Freon volume chosen initially, 5 in.³, is adequate to expel all the liquid mercury. Second, the bladder and tank design is satisfactory and high expulsion efficiencies (99.9%) are obtained.

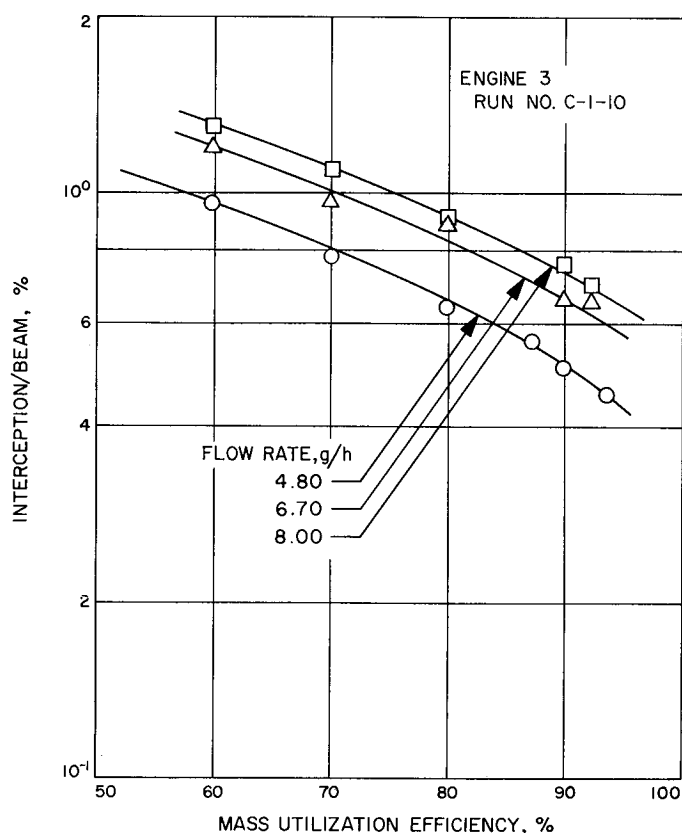


Fig. 21. Effect of flow rate and mass utilization on interception

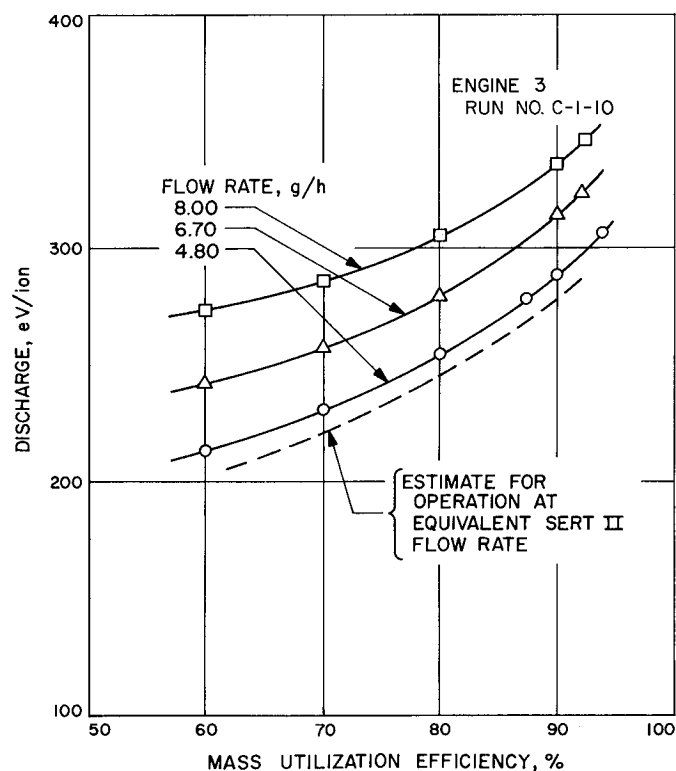


Fig. 22. Present system efficiency

References

1. Kerrisk, D. J., and Kaufman, H. R., "Electric Propulsion Systems for Primary Spacecraft Propulsion," AIAA Paper 67-424, AIAA Third Propulsion Joint Specialists Conference, Washington, D.C., July 1967.
2. Flandro, G. A., and Barber, T. A., "Low Thrust Mission Analysis," AIAA Paper 67-678, AIAA Electric Propulsion and Plasmadynamics Conference, Colorado Springs, Colo., Sept. 1967.
3. Sauer, C. G., Jr., "Trajectory Analysis and Optimization of a Low-Thrust Solar-Electric Jupiter Flyby Mission," AIAA Paper 67-680, AIAA Electric Propulsion and Plasmadynamics Conference, Colorado Springs, Colo., Sept. 1967.
4. Masek, T. D., and Womack, J. R., "Experimental Studies with a Clustered Ion Engine System," AIAA Paper 67-698, AIAA Electric Propulsion and Plasmadynamics Conference, Colorado Springs, Colo., Sept. 1967.
5. Bechtel, R. T., "Discharge Chamber Optimization of the SERT II Thruster," AIAA Paper 67-668, AIAA Electric Propulsion and Plasmadynamics Conference, Colorado Springs, Colo., Sept. 1967.

XIII. Liquid Propulsion

PROPULSION DIVISION

A. The Liquid-Phase Mixing of a Pair of Impinging Sheets, R. W. Riebling

1. Introduction

When a jet of liquid of diameter d is directed tangentially against a solid deflector surface of radius R and angle θ , it spreads to form a thin liquid sheet of width w (Fig. 1). Upon leaving the deflector, the sheet spreads through an angle β and deflects slightly through an angle δ before finally breaking up into droplets. The manner in which the dimensions and spatial orientation of such liquid sheets, as well as their mass flux, momentum flux, velocity and thickness distributions, vary with deflector geometry, injection velocity, and propellant physical properties was reported earlier (Ref. 1).

Injector elements incorporating this effect are being developed at JPL because they offer certain advantages over the more conventional impinging-jet varieties. One such device is the unlike impinging-sheet injector element, shown schematically in Fig. 2, which brings flat sheets of fuel and oxidizer together along an impingement line,

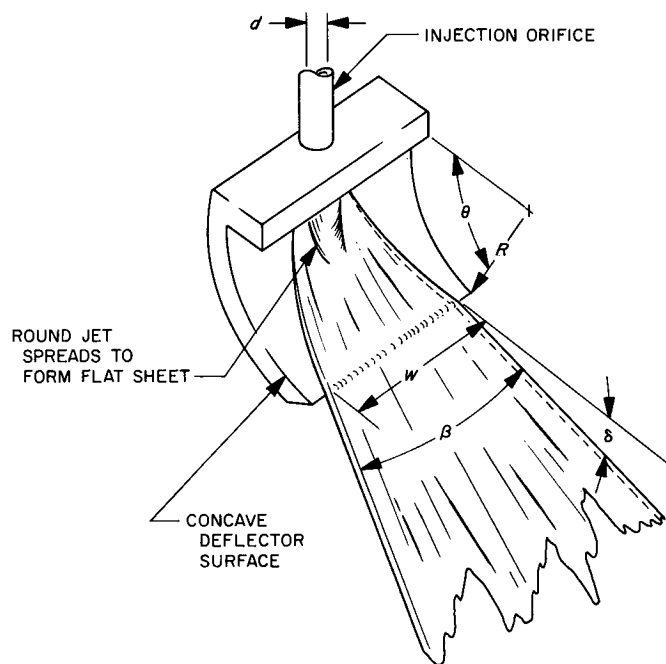


Fig. 1. Sheet formation on a deflector

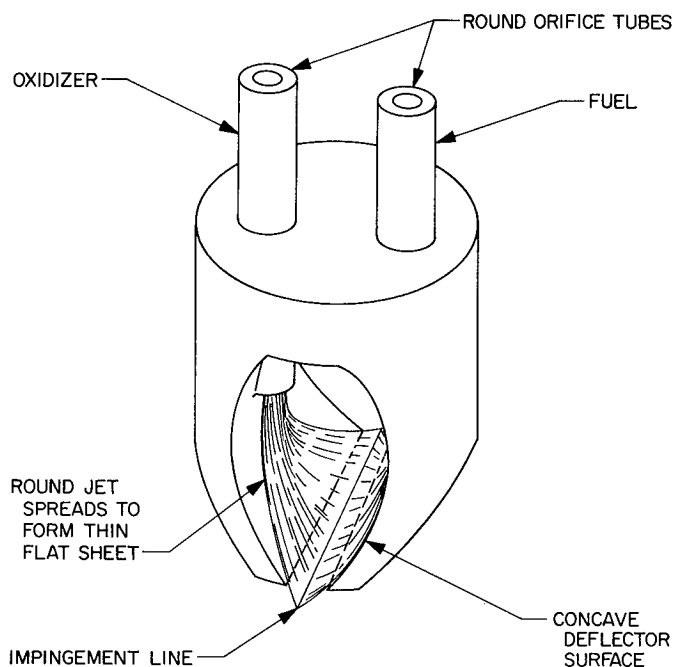


Fig. 2. Typical impinging-sheet injector element

where a finely atomized spray is formed. Results of recent firings of this type of injector are reported in SPS 37-41, Vol. V; SPS 37-44, Vols. IV and V; and SPS 37-48, Vol. IV.

A brief experimental study was conducted to determine the effect of several key injector variables on the degree of primary, liquid-phase mixing attainable in the sprays from single, unlike impinging-sheet elements. This report will present and discuss the results of that investigation.

2. Apparatus

Four small impinging-sheet elements were constructed of stainless steel. Each element consisted of a pair of identical deflectors and orifices, the key dimensions of which are summarized in Table 1. It should be noted that no two elements had the same combination of R , d , and θ . The overhang ratio h/d included in Table 1 for each

Table 1. Dimensions of impinging-sheet elements

Element no.	Orifice diameter d , in.	Deflector radius R , in.	Deflector angle, θ , deg	Overhang ratio, h/d
1	0.020	0.322	30	2.16
2	0.020	0.147	45	2.16
3	0.041	0.147	45	1.06
4	0.032	0.147	45	1.37

element is a parameter found quite useful for correlating the properties of *single* liquid sheets (Ref. 1). It is derived from the deflector overhang

$$h = R(1 - \cos \theta) \quad (1)$$

which is the transverse distance to which the deflector protrudes into the otherwise undisturbed round jet. Elements 1 and 2 were intentionally designed to have identical values of h/d , even though their other dimensions were quite different. All elements were constructed so that the spacing between the edges of their deflectors could be continuously varied from 0.1 to 0.2 in.

Each element was mounted in a test fixture and sprayed vertically into the JPL spray collector (Fig. 3) from a fixed height of 3.5 in. The mounting frame could be both translated and tilted so that the spray axis was maintained vertical and aligned with the center of the collection rake in each test. The inert, immiscible propellant simulants used were water and trichlorethylene, simulating hydrazine (N_2H_4) and nitrogen tetroxide (N_2O_4), respectively.

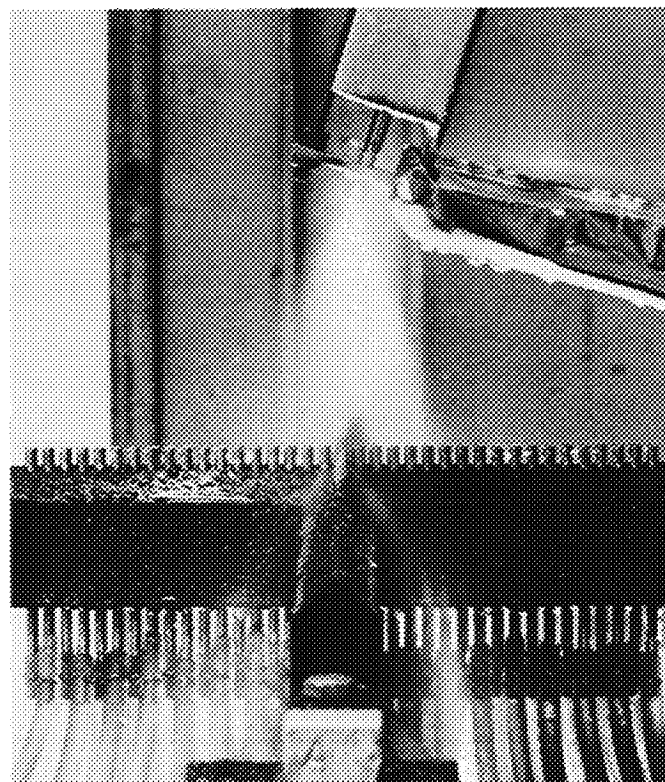


Fig. 3. Impinging-sheet injector and spray collector during flow test with nonreactive propellant simulants

By sampling with the collector rotated at various angles, the distributions of mass flux and mixture ratio in a transverse plane across the whole spray were determined.

3. Test Conditions

Each element was flowed at weight-flow mixture ratios (trichlorethylene to water) ranging between 0.7 and 2.0. The corresponding stream momentum flux ratios (trichlorethylene to water) varied from 0.34 to 2.76. Injection velocities varied between 25 and 125 ft/s.

4. Results and Discussion

Typical contour maps of the mass flux and mixture ratio distributions obtained with the impinging-sheet elements are shown in Figs. 4 and 5. Geometrically, the sheet-spray patterns are quite similar to those found for unlike doublet jet elements. That is, elongated crescent- or elliptical-shaped spray fans are produced, with the mass flux being symmetrically distributed about a maximum point near the center. There is typically a steady gradient in simulant mixture ratio across the spray from fuel-rich conditions on the oxidizer orifice side to oxidizer-rich conditions on the fuel orifice side. This indicates that considerable interpenetration of the two sheets occurs under nonreactive conditions.

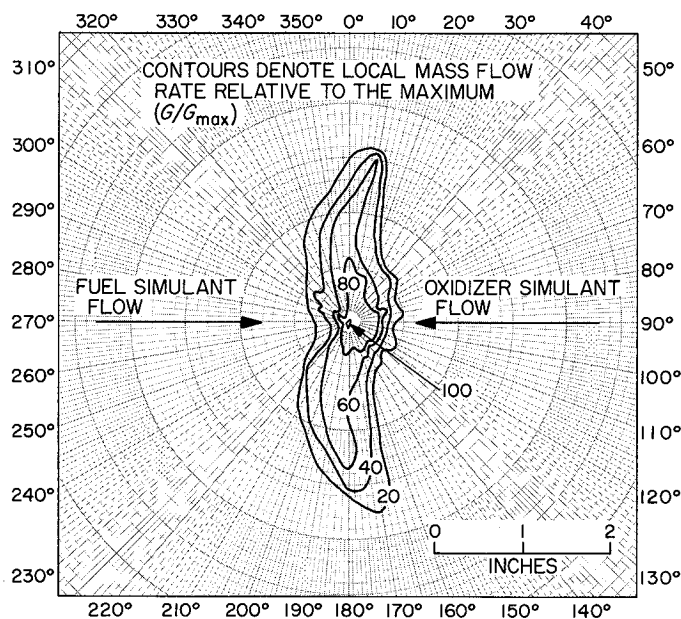


Fig. 4. Typical mass flux distribution in spray from single-element impinging-sheet injector using nonreactive propellant simulants

From a knowledge of local values of mass flux and mixture ratio all across the sprays, it was possible to calculate values of the mixing factor, E_m . The mixing factor, first introduced in Ref. 2, is a measure of the degree to which the input mixture ratio to the injector is attained locally within the spray. For perfect mixing, E_m would equal 100.

The experimental results, in terms of E_m , are plotted in Figs. 6 and 7. Fig. 6 shows the effects of deflector overhang ratio h/d and a momentum flux and diameter ratio grouping R on E_m . R is defined as

$$R = \frac{1}{1 + \phi} \quad (2)$$

where ϕ is the stream momentum flux ratio divided by the orifice diameter ratio

$$\phi = \frac{m_f/m_o}{d_f/d_o} \quad (3)$$

First introduced in Ref. 3 as

$$\phi = \frac{\rho_f V_f^2 d_f}{\rho_o V_o^2 d_o} \quad (4)$$

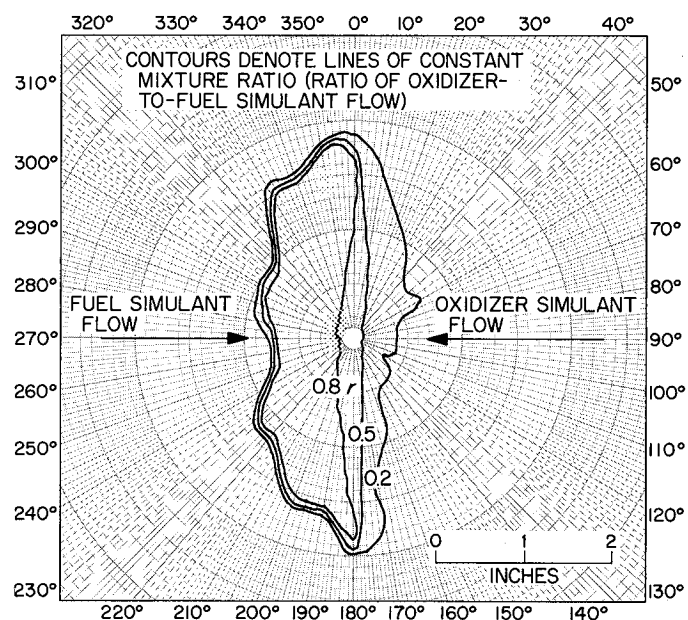


Fig. 5. Typical mixture ratio distribution in spray from single-element impinging-sheet injector using nonreactive propellant simulants

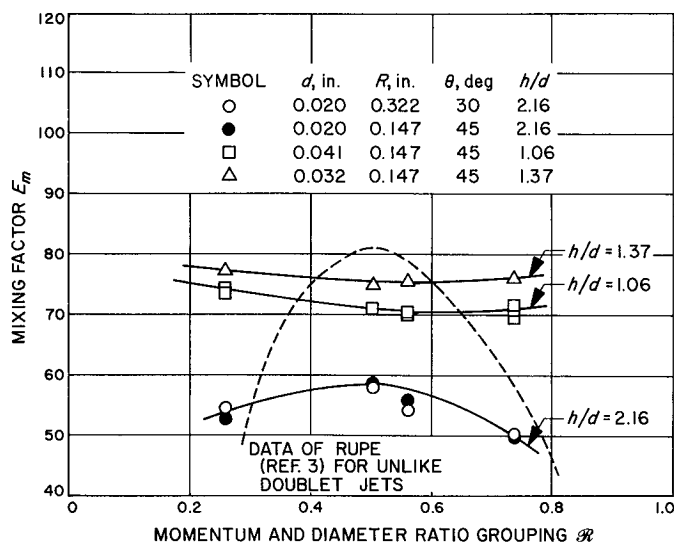


Fig. 6. Effect of momentum and overhang ratios on mixing factor at a constant spacing of 0.1 in.

it may also be expressed in the equivalent form

$$\phi = \left(\frac{1}{r}\right)^2 \left(\frac{\rho_o}{\rho_f}\right) \left(\frac{d_o}{d_f}\right)^3 \quad (5)$$

\mathcal{R} is therefore a function of liquid physical properties (ρ_f/ρ_o), injection velocity (V_f/V_o)², and injector geometry (d_f/d_o). It has been found to be quite useful in correlating E_m for a wide variety of unlike doublet jet elements (Ref. 3).

The results plotted in Fig. 6 indicate that for a pair of unlike doublet sheets, E_m is relatively insensitive to \mathcal{R} . This is in contrast to the behavior commonly observed with unlike doublet jet elements, which characteristically exhibit a sharp E_m peak at $\mathcal{R} = 0.5$. For purposes of comparison, Rupe's curve for unlike impinging jets from Ref. 3 is superimposed on Fig. 6. The relative flatness of the sheet curves suggests that there may be no optimum momentum ratio for sheets as there is for jets; rather, wide excursions in mixture ratio r (and therefore in \mathcal{R}) may be possible before the degree of mixing is seriously degraded. It should be pointed out that for the experiments reported herein, the density and diameter ratios were held constant, and \mathcal{R} was varied only by adjusting injection velocities. Future experiments will vary \mathcal{R} by changing fluid properties and injector geometry. The verification of \mathcal{R} as a true correlating parameter for E_m with sheet elements must await the results of those tests.

Fig. 6 also reveals that the overhang ratio h/d has a major effect on E_m , with $h/d = 1.37$ giving the highest

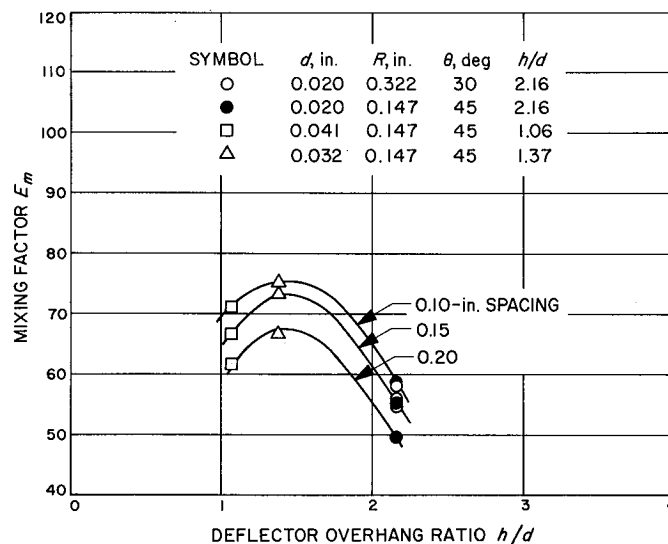


Fig. 7. Effect of deflector spacing and overhang ratio on mixing factor at $\mathcal{R} = 0.5$

values. The value of h/d below this (1.06) and the value above it (2.16), both yielded less efficient mixing. These results suggest that there is some optimum value of h/d for maximum E_m , although it cannot be too closely pinpointed at this time because only three values of h/d were investigated. It probably falls within the range $1.3 \leq h/d \leq 1.6$.

The influence of h/d on E_m is not yet completely understood, although it is not too surprising, considering the first-order effects of h/d on most sheet dimensions and on the distributions of mass flux, velocity, momentum flux and thickness across the sheets (Ref. 1).

That the significant variable is h/d rather than any of the individual dimensions is substantiated by the data for $h/d = 2.16$, obtained with elements 1 and 2 (Table 1), which had different values of r , d and θ .

The data points on Fig. 6 represent tests conducted at injection velocities ranging from 25 to 125 ft/s; it is seen that it is the ratio of these velocities (as it influences \mathcal{R}) and not their absolute values, which exerts the primary influence on E_m . However, injection velocity levels may affect the degree of atomization attained in the spray, as they do with impinging jets.

Another geometric variable, the deflector spacing, has a major effect on E_m , as shown in Fig. 7, at a constant value of \mathcal{R} (0.5). At each value of h/d , E_m decreases as the spacing is made wider. The spacing, of course,

Definition of terms

d = orifice diameter, in.
 G = mass flux per unit area, lb/s/in.²
 m = stream momentum flux, ft-lbm/s²
 R = deflector radius, in.
 r = mixture ratio, oxidizer simulant/fuel simulant
 V = injection velocity, ft/s
 ρ = liquid density, lbm/ft³

Subscripts

o = oxidizer simulant
 f = fuel simulant

affects the free-sheet length before impingement, and as this length is increased, the velocity profiles initially developed while the sheet was in contact with the deflector surface will decay, tending to become more nearly uniform. In addition, the thin sheets begin breaking up

into ligaments and droplets soon after leaving the deflectors; the greater the free sheet length, the greater the degree to which this breakup will have occurred before impingement. The stream momentum can better be utilized for the purposes of mixing by impacting integral, rather than partially broken-up, sheets.

Figure 7 also shows the key influence of h/d on E_m at each of the three spacings, and indicates that maximum E_m probably occurs somewhere around $h/d \cong 1.5$.

References

1. Riebling, R. W., *The Formation and Properties of Liquid Sheets Suitable for Use in Rocket Engine Injectors*, Technical Report 32-1112, Jet Propulsion Laboratory, Pasadena, Calif., June 15, 1967.
2. Rupe, J. H., *The Liquid-Phase Mixing of a Pair of Impinging Streams*, Progress Report 20-195, Jet Propulsion Laboratory, Pasadena, Calif., Aug. 6, 1953.
3. Rupe, J. H., *A Correlation Between the Dynamic Properties of a Pair of Impinging Streams and the Uniformity of Mixture Ratio Distribution in the Resulting Spray*, Progress Report 20-209, Jet Propulsion Laboratory, Pasadena, Calif., Mar. 28, 1956.

PRECEDING PAGE BLANK NOT FILMED.

XIV. Space Instruments

SPACE SCIENCES DIVISION

A. Sterilizable, Ruggedized Imaging System,

L. R. Baker

1. Introduction

To undertake scientific observations on the surface of Mars, a television instrument can perform visual observations to observe distinct local topographic variations, and perhaps observe experiments in process.

The instrument design must meet specific requirements such as required by planetary quarantine requirements and possible high g impact upon landing. A television system has definite advantages for performing experiments on the surface.

2. Program Outline

A program to develop a TV system capable of landing on Mars has been under way since May 1965. The primary problems in meeting sterilization and ruggedization are directly connected with the image sensor. The packaging and associated electronics are a straightforward application of well-known component selection, careful electronic packaging, and thoughtful circuit design.

The program was divided into two phases. Phase I was to develop an operational breadboard television system using the *Mariner* Mars 1964 television system as a generic system with the operating parameters and system design as shown in Table 1.

Table 1. Ruggedized sterilizable system design

Scan lines per frame	512
Pixels per line	512
Frame time	13.65 s
Active line time	25 ms
Beam chopper frequency	76.8 kHz
Video baseband	10.24 kHz
Vidicon type	TCA-C23086
Focus	Electro-static
Deflection	Magnetic
Scanned area	11 mm square
Timing	Digital, synchronized
Circuitry	Functionally integrated
Encoding rate	1.54 mHz, 6-bit RzPCM
Scanning	Analog
Volume	≈ 400 in. ³
Weight	≈ 9 lb
Power	≈ 8.25 W

The objectives of Phase I were to design and fabricate a breadboard model of a slow-scan television system which would survive sterilization of 6 cycles of 76 hr each at 135°C, and survive ethylene oxide decontamination of 6 cycles of 28 hr each at 40°C. The criteria used in the breadboard system design were the following:

- (1) No specific mission; design the system as a simple low-power television camera
- (2) Self-contained instrument with digital data output
- (3) Low data rate output
- (4) Use of a *Mariner* Mars 1964-type shutter
- (5) Ease of operation for evaluation purposes
- (6) $\pm 1/2$ DN encoding accuracy
- (7) 2% regulation power supply to conserve power
- (8) No engineering telemetry
- (9) All timing functions to be internal
- (10) Use of a simple gain control computer

Phase II consists of the fabrication and sterilization testing of an engineering model of the imaging system using the circuit design from Phase I.

3. Image Sensor

As mentioned, the primary problem in developing a sterilizable television system is obtaining an image sensor which can be sterilized. Until recently, the technology related to vidicons for space imaging system application did not produce vidicons capable of meeting present sterilization requirements. Additionally, it can be assumed that any sterilizable component might find application in a high-impact environment. Therefore, a program was conceived to design, fabricate, and test a vidicon-type image sensor capable of surviving dry heat sterilization and ethylene oxide decontamination. The vidicon elements and structure are being designed to withstand a 3000-g shock. The results of the vidicon program were reported in SPS 37-43, Vol. IV, pp. 264-273.

4. Results

The breadboard operated satisfactorily, but problems with the C23086 vidicon forced the use of a *Mariner* Mars 1964 vidicon. The sterilizable vidicon has a very high dark current, and insufficient time was available to make the required circuit changes and complete the breadboard evaluation. Briefly, the following results were obtained: signal to noise ratio of 39 dB at 0.1 ft-cd-s exposure; reso-

lution of 58% at 200 TV lines and 17% at 400 TV lines; deviation of the digital signal to the analog signal of less than 1%; system power consumption of less than 7 W.

B. Photo Sensor Evaluation, K. J. Ando and L. R. Baker

1. Introduction

Imaging system development for lunar and planetary space programs requires extensive test and evaluation of image sensor performance in terms of those characteristics which define image quality. Evaluation of image and optical parameters requires very specialized techniques and equipment. Figures 1 and 2 show part of the equipment necessary. Not shown is a Bemco thermal vacuum chamber which is equipped with a 20-in.-diam optically flat fused-silica window.

The image sensor test set shown in Fig. 1 has been designed to perform image quality tests on various types of image sensors such as would be used for lunar and planetary exploration space programs. The test set can operate at any scan rate between standard TV scan rates (EIA, 1/30th s/frame) to very slow scan rates of 30 s/frame. Additionally, a mechanical shutter is available for slow-scan shuttered operation tests.

Image quality tests that the equipment can perform are: light transfer, characteristic of luminance input versus target current output, dark current, resolution, coherent noise, sensitivity, image storage, residual image, field uniformity, and spectral sensitivity.

2. Current Activities

The test set has been used quite extensively to evaluate developmental, prototype, and flight image sensors for programs such as *Ranger*, *Mariner IV*, *Surveyor*, and *Mariner* Mars 1969. The work on evaluating advanced image sensors has included sterilizable vidicons, SEC vidicons, and return-beam vidicons. SECVs and RBVs are relatively new image sensors which appear to exhibit significant improvement in performance relative to the usual 1-in. vidicon. It is the purpose of this evaluation program to determine if further image sensor development is necessary or if current sensors can have any application in the planetary exploration program. Figure 3 shows the SECV mount which houses a type WX 30691 SECV, and Fig. 4 shows the RBV mount, still under construction, which will house a type C23061A 2-in. RBV. Evaluation on both the above devices constitutes the primary effort in this program for FY 68.

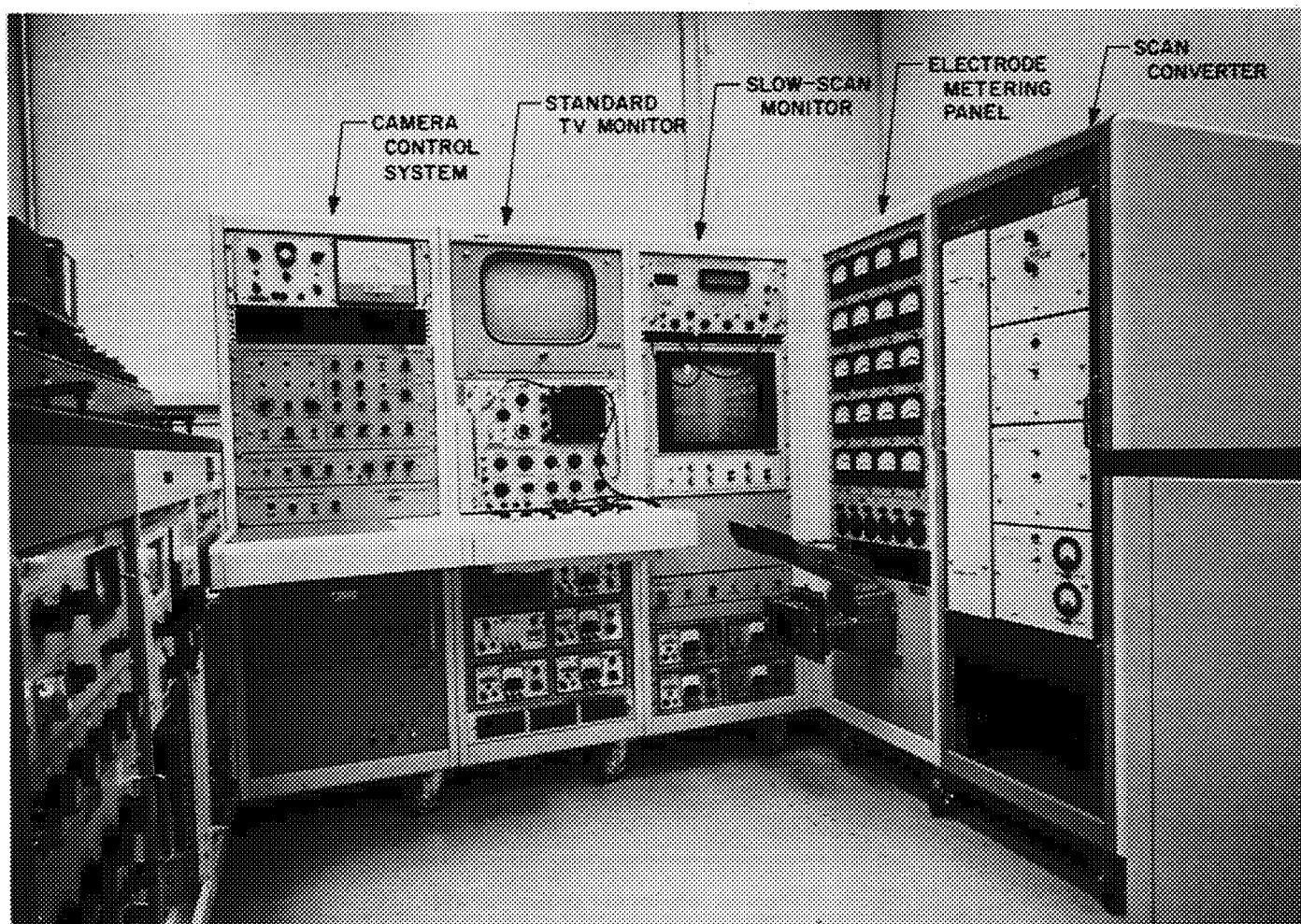


Fig. 1. Image sensor test set

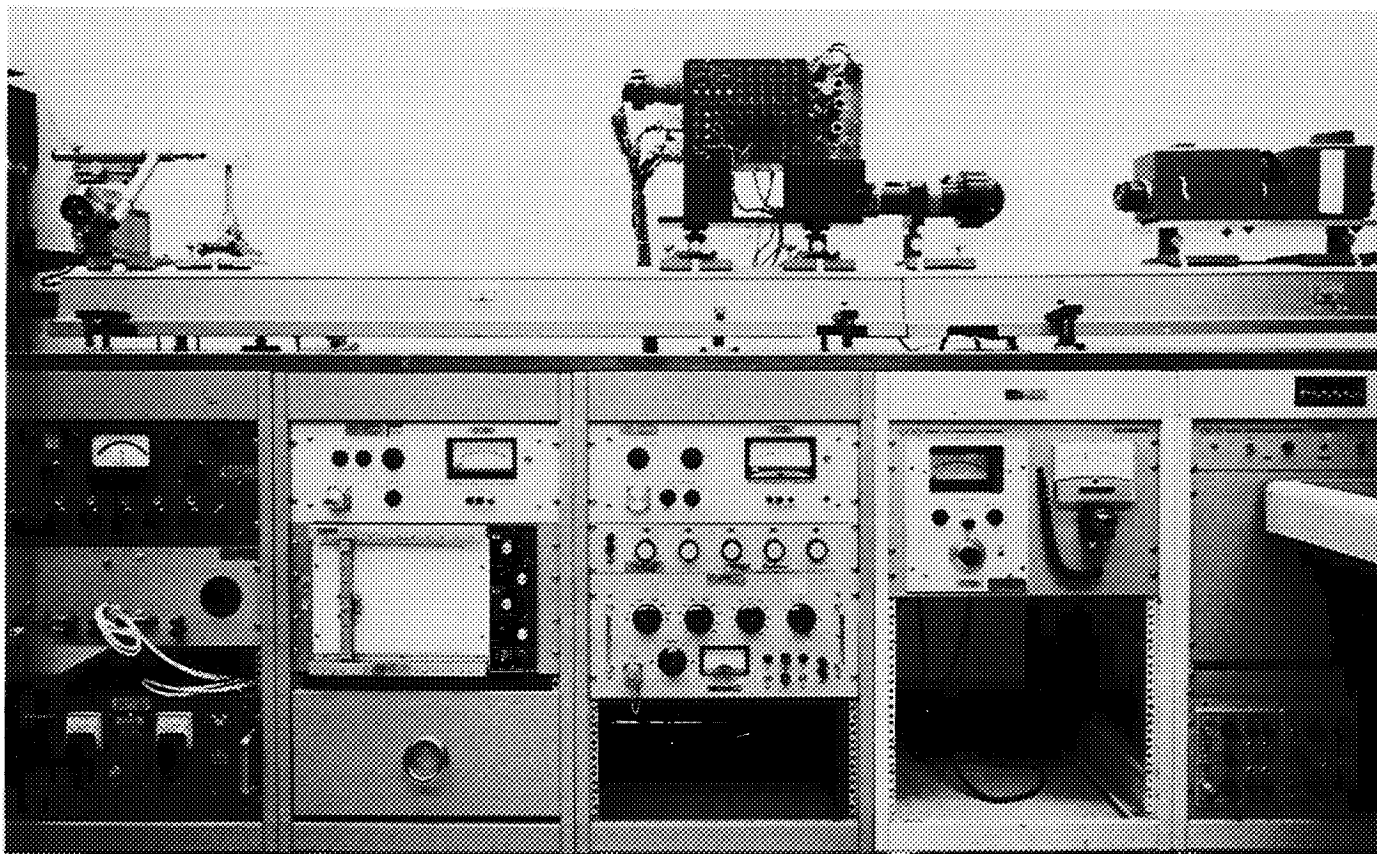


Fig. 2. Optical bench, image sensor test set

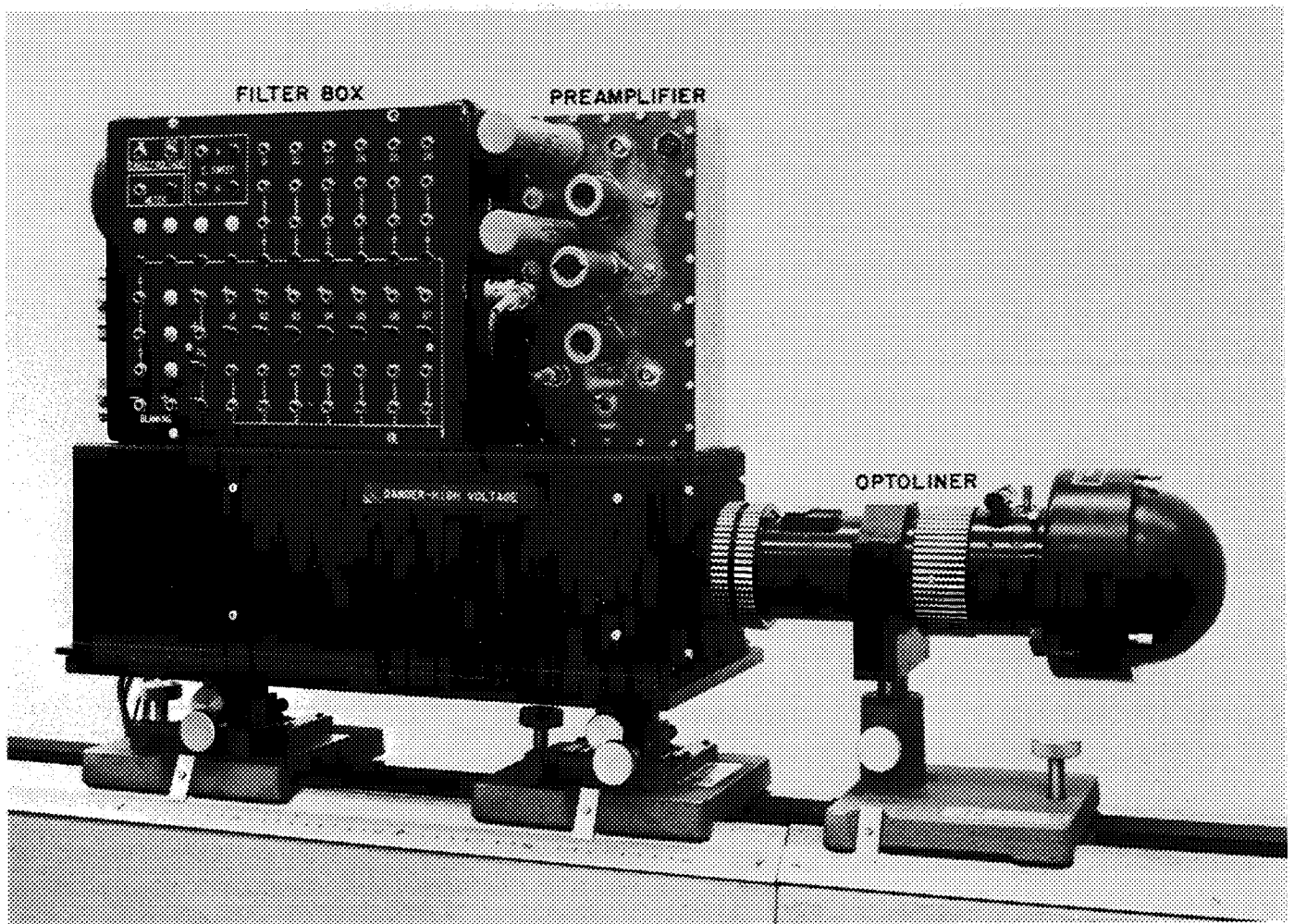


Fig. 3. SEC vidicon mount

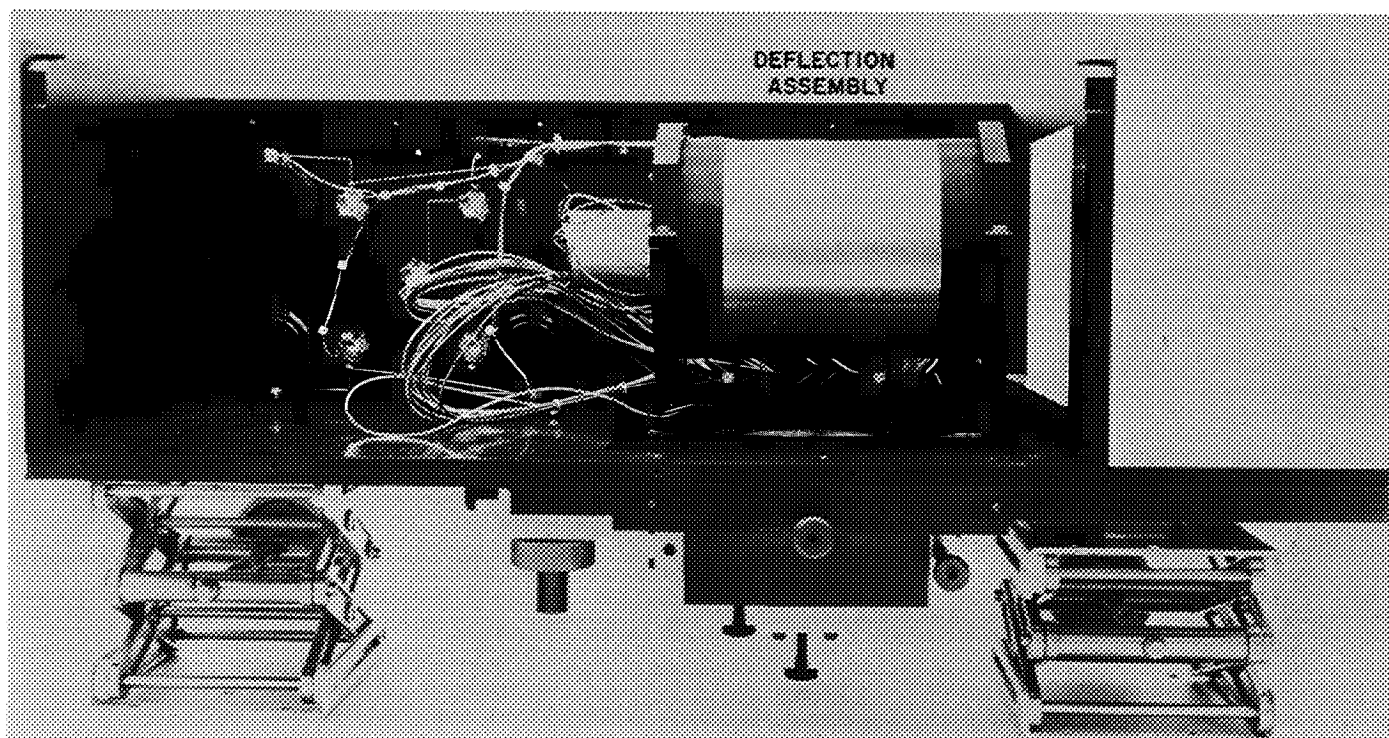


Fig. 4. RBV mount

In addition to the image sensor evaluations, work is progressing to improve the test set performance and to update the instrumentation associated with the test set. Modifications and additions to the test set which are under way are: (1) Increase sweep current drive, and linearity to allow the evaluation of large format image sensors; and (2) improve the video amplification and processing with respect to noise, dynamic range, gain, phase response, distortion, and bandwidth.

C. Studies on the Photoconducting Layer in Slow-Scan Vidicons, K. J. Ando and L. R. Baker

1. Introduction

As a program to supplement the testing and evaluation of vidicons and other imaging devices for *Mariner* Mars 1969 and future missions, a study has been initiated to understand the solid state properties of the photoconducting layer in vidicons operating in a slow-scan mode.

The vidicon is a television camera tube employing a thin layer of photoconducting material as its light-sensitive element. The spectra sensitivity, light transfer, storage time, and erase characteristics of a vidicon depend primarily on the properties of this layer.

The purpose of this study is to define more clearly the physics of the photoconducting layer and develop that information which will aid in understanding the effects of the properties of the photoconducting layer on vidicon performance. The study will be performed in two phases:

a. Phase I. In this phase, information and data on photoconductivity and vidicons will be obtained through a detailed literature survey, through contacts and visits to manufacturers and research organizations, and participation in technical meetings on imaging devices. This information will be utilized to determine the current state of the art in vidicons and other imaging devices. In addition, a physical model describing the behavior of the photoconducting layer and its effects on vidicon performance characteristics will be formulated.

b. Phase II. In this phase, vidicon parameters which are normally not determined in a performance evaluation but which can yield information concerning the nature of the photoconducting layer will be experimentally measured.

A complete summation of the results of Phases I and II will be reported in a JPL Technical Report now in preparation.

Studies to date indicate that the vidicon read cycle can be discussed on the basis of a simple equivalent circuit. Light incident on the photoconducting layer of a vidicon forms a charge image which is read out by a low velocity electron beam scan resulting in a variation of the beam current. This variation constitutes the output current. In previous equivalent circuits of a vidicon, (Ref. 1), the target elements are assumed to charge up through an effective beam impedance R_B and a target load resistor R_L as shown in Fig. 5. In the present case, R_B and R_L are taken into consideration by assuming a theoretical beam acceptance current given by (Ref. 2)

$$\begin{aligned} i &= i_0 & \text{for } V \geq 0 \\ i &= i_0 \exp(eV/kT_0) & \text{for } V \leq 0 \end{aligned} \quad (1)$$

where e is the electronic charge, k is Boltzmann's constant, and T_0 is the effective cathode temperature. If we write $V_0 = kT_0/e$, Eq. (1) can be written

$$\begin{aligned} i &= i_0 & \text{for } V \geq 0 \\ i &= i_0 \exp(V/V_0) & \text{for } V \leq 0 \end{aligned} \quad (2)$$

If the capacitor C in Fig. 6 is now charged by the beam current given by Eq. (2), the potential V of the capacitor as a function of time can be calculated by integrating the differential equation:

$$-idt = cdV \quad (3)$$

For $V \geq 0$ and $t \leq 0$, this gives

$$V = -i_0 t/C \quad (4)$$

if $V = 0$ at $t = 0$.

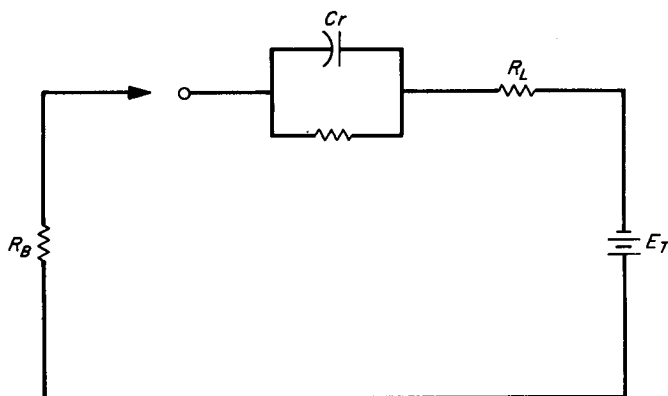
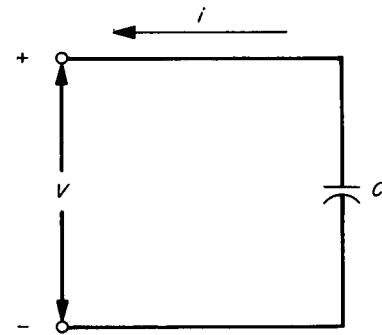


Fig. 5. Equivalent circuit of a vidicon target element



C = CAPACITANCE OF THE TARGET ELEMENT
 V = POTENTIAL OF THE TARGET ELEMENT
 i = INCIDENT BEAM CURRENT

Fig. 6. Charging of a target element

For $V \leq 0$, we find

$$V/V_0 = -\log(1 + i_0 t/CV_0) \quad (4a)$$

again with $V = 0$ at $t = 0$.

Introducing a dimensionless time parameter

$$x = i_0 t/CV_0$$

Equations (4) and (4a) can be rewritten as

$$\begin{aligned} V/V_0 &= -x & (V \geq 0) \\ V/V_0 &= -\log(1 + x) & (V \leq 0) \end{aligned} \quad (5)$$

Equation (5) is plotted in Fig. 7 with the cathode at $V = 0$. It shows how the potential of the capacitor drops

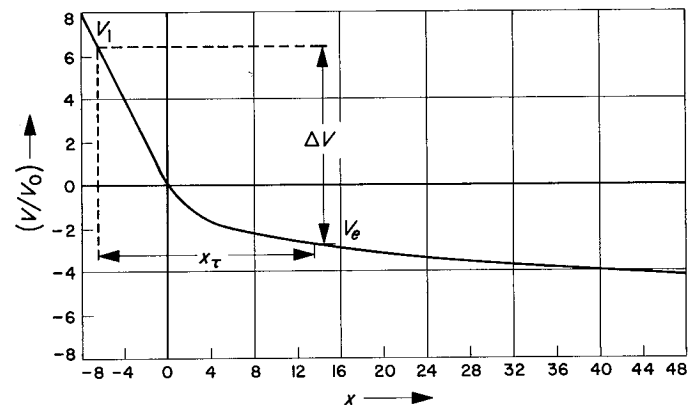


Fig. 7. Potential of target element as a function of time for theoretical beam current given by Eqs. (4) and (4a)

as a function of the charging time. the initial charging rate is determined by the values of the constants i_0 , C and V_0 .

If the time that the beam is incident on the target element and the initial voltage V_1 is known, the end voltage V_e can be determined from Fig. 7. Thus, V_1 is the potential just before scanning and V_e is the potential just after scanning. The time interval x_τ in Fig. 7 is given by $x_\tau = i_0 \tau / CV_0$ where τ is the time during which the beam current flows. The voltage drop ΔV over the target element is given by

$$\Delta V = V_1 - V_e \quad (6)$$

and the average signal current i_s during the time interval τ is

$$i_s = C\Delta V/\tau \quad (7)$$

From Eq. (7) it follows that if ΔV is known, the average output current can be determined. Suppose that the capacitor is at some positive potential V_1 and then charged to a negative potential V_e during the read scan by the electron beam. The average signal current can be

written in terms of x_1 and x_e as

$$i_s = \frac{i_0(\log(x_1 + 1) - x_e)}{(-x_1 + x_e)} \quad (8)$$

For $x_e \gg x_1$, i_s decreases as x_e increases, with $i_s \propto 1/\tau$. Thus it follows that the signal current decreases as the scan rate is decreased. Experimental measurements of the beam acceptance currents (Eq. 2), in vidicons agree well with the beam acceptance curves given by Eqs. (4) and (4a) except that there is a shift of the beam acceptance curves toward higher potentials due to the contact potential between the photoconducting layer and signal electrode.

If the target capacitance and the initial beam impedance limited current is known, the preceding analysis can be utilized to determine the average current expected during slow scan operation. During Phase II, the target capacitance and other parameters of the slow scan type vidicons will be measured and a comparison will be made with the results above.

References

1. Malling, L. R., *J. of SMPTE*, Vol. 72, p. 872, 1963.
2. van de Polder, L. J., *Philips Research Reports*, Vol. 22, p. 178, 1967.

XV. Science Data Systems

SPACE SCIENCES DIVISION

A. Piece-wise Linear Approximation of a Mass Spectrometer Sweep Voltage, W. Spaniol

1. Introduction

In a mass spectrometer, the mass number of a detected element is proportional to the instantaneous sweep voltage. This voltage ranges as high as 1000–2000 V and is difficult to measure with high precision without using large amounts of power. If the voltage were generated with high precision by a time-related function, an accurate measure of time would precisely define the voltage. A commonly used function is the exponential discharge of a resistor and capacitor circuit. This type of circuit is extremely sensitive to changes in values of the components. A digital approach to function generation seems to offer an ideal solution.

This article deals with digital generation of a particular function, $V = 1800 \exp(-0.44t)$, which would be used as a mass spectrometer sweep voltage. A linear approximation approach was chosen and a criterion for segmenting was developed. The complete logical design was performed and a breadboard model was built and tested.

The generator will be integrated into the mass spectrometer, and an evaluation of the instrument system will be made to determine the performance of this approximation. The number of segments will be increased or decreased until an optimum approximation is defined.

2. Design Development

The curve to be approximated is $V = 1800 \exp(-0.44t)$ over the range $V = 1800$ V to $V = 200$ V, $t = 0$ to $t = 5$ s. For the breadboard, the following criteria were used in making the approximations:

Consider the portion of the sweep voltage shown in Fig. 1. At P_1 , the slope S_1 of the exponential curve is greater than the slope of the straight line joining P_1 and P_2 . At P_2 , the slope of the line is greater than the slope of the curve S_2 . For small values of Δt , the three slopes are very nearly equal. The criteria used to determine Δt are

$$k = \frac{S_1}{M_L} = \frac{M_L}{S_2} \quad (k \text{ approaches } 1 \text{ as } \Delta t \text{ approaches } 0)$$

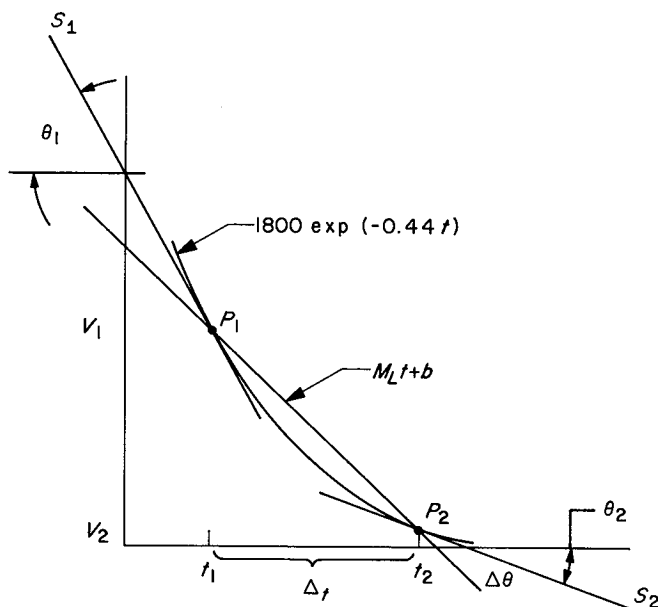


Fig. 1. A general segment of the exponential curve

where, at P_1 ,

$$S_1 = \frac{dV_1}{dt} = (-0.44)(1800) \exp(-0.44t_1) = -792 \exp(-0.44t_1)$$

and, at P_2 ,

$$S_2 = \frac{dV_2}{dt} = (-0.44)(1800) \exp(-0.44t_2) = -792 \exp(-0.44t_2)$$

and where

$$M_L = \frac{V_2 - V_1}{t_2 - t_1} = \frac{1800 [\exp(-0.44t_2) - \exp(-0.44t_1)]}{t_2 - t_1}$$

Table 1. Relative improvement in k as the number of segments is increased

n	No. of segments	t, s	$0.44\Delta t$	$\exp(-0.44\Delta t)$	$1 - \exp(-0.44\Delta t)$	k
0	1	5.0	2.2	0.110	0.890	2.47
1	2	2.5	1.1	0.333	0.667	1.65
2	4	1.25	0.55	0.577	0.423	1.3
3	8	0.625	0.275	0.760	0.240	1.145
4	16	0.3125	0.1375	0.873	0.127	1.084
5	32	0.15625	0.06875	0.9335	0.0665	1.033
6	64	0.078125	0.034375	0.9663	0.0337	1.018

Since

$$t_2 - t_1 = \Delta t$$

it follows that

$$M_L = \frac{1800}{\Delta t} \exp(-0.44t_1) [\exp(-0.44\Delta t) - 1]$$

$$k = \frac{S_1}{M_L} = \frac{(-0.44)(1800) \exp(-0.44t_1)}{\frac{1800}{\Delta t} \exp(-0.44t_1) [\exp(-0.44\Delta t) - 1]}$$

$$= \frac{0.44\Delta t}{1 - \exp(-0.44\Delta t)}$$

Therefore, k can be shown to be constant throughout the sweep if the time increments Δt are equal, approaching 1 as Δt approaches 0. The implementation is simplified if the time base is divided by powers of 2, i.e., $\Delta t = 5/2^n$. Table 1 lists values obtained by substituting 0 to 6 for n . Obviously the approximation improves as n increases, and the complexity of the logic implementation also increases. For the breadboard, $n = 4$ was chosen as the best compromise.

The following calculations describe, as functions of n , the angle $\Delta\theta$ between the tangent to the exponential and the straight line segment, at the points of intersection of the curve and the line; and dV/V , the ratio of slope to voltage:

$$k = \frac{S_1}{M_1} = \frac{\tan \theta_1}{\tan \theta_2}$$

$$= \frac{\sin \theta_1 \cos \theta_2}{\cos \theta_1 \sin \theta_2}$$

$$- k \cos \theta_1 \sin \theta_2 + \sin \theta_1 \cos \theta_2$$

$$= 0$$

Subtracting $(1 - k) \cos \theta_1 \sin \theta_2$ from each side and applying the appropriate trigonometric identity yields

$$\sin(\theta_1 - \theta_2) = (k - 1) \cos \theta_1 \sin \theta_2$$

Since θ_1 very nearly equals θ_2 , it can be shown that $\cos \theta_1 \times \sin \theta_2$ has a maximum value of 0.5:

$$\theta_1 - \theta_2 = \Delta\theta = \sin^{-1}[(k - 1) \cdot 0.5]$$

Intermediate values of n are given in Table 2.

Table 2. Intermediate values of n

n	$\Delta\theta$, deg
3	4.02
4	2.41
5	0.945
6	0.504

For the exponential curve, the ratio of slope to voltage is constant:

$$V = 1800 \exp(-0.44t)$$

$$dV = (-0.44)(1800) \exp(-0.44t)$$

$$\frac{dV}{V} = -0.44$$

For the straight-line approximation,

$$V = M_L t + b$$

$$dV = M_L$$

At $t = t_1$, the curve and the line intersect, making it possible to solve for the constant b :

$$V = 1800 \exp(-0.44t_1) - M_L t_1 + b$$

$$b = 1800 \exp(-0.44t_1) - \frac{1800}{\Delta t} \exp(-0.44t_1) [\exp(-0.44\Delta t) - 1] t_1$$

$$= 1800 \exp(-0.44t_1) \left\{ \frac{\Delta t - [\exp(-0.44\Delta t) - 1] t_1}{\Delta t} \right\}$$

$$V = M_L t + b$$

$$= \frac{1800 \exp(-0.44t_1)}{\Delta t} \{ [\exp(-0.44\Delta t) - 1] (t - t_1) + \Delta t \}$$

$$\frac{dV}{V} = \frac{M_L}{M_L t + b} = \frac{[\exp(-0.44\Delta t) - 1]}{[\exp(-0.44\Delta t) - 1] (t - t_1) + \Delta t}$$

where Δt is defined as $t/2^n$. Table 3 lists values of dV/V as a function of n . These data are shown graphically in Fig. 2.

Figure 3 shows the actual rate of change of voltage for the exponential and the approximation. The staircase effect is due to the constant slope of the straight-line segments.

Table 3. Values of dV/V as a function of n

n	Δt	exp (-0.44 Δt)	$\Delta t \exp$ (-0.44 Δt)	dV/V	
				Minimum	Maximum
0	5.0	0.110	0.550	0.176	1.615
1	2.5	0.333	0.835	0.267	0.800
2	1.25	0.577	0.723	0.338	0.585
3	0.625	0.760	0.475	0.384	0.506
4	0.3125	0.873	0.272	0.407	0.467
5	0.15625	0.9335	0.146	0.425	0.448
6	0.078125	0.9663	0.0755	0.431	0.446

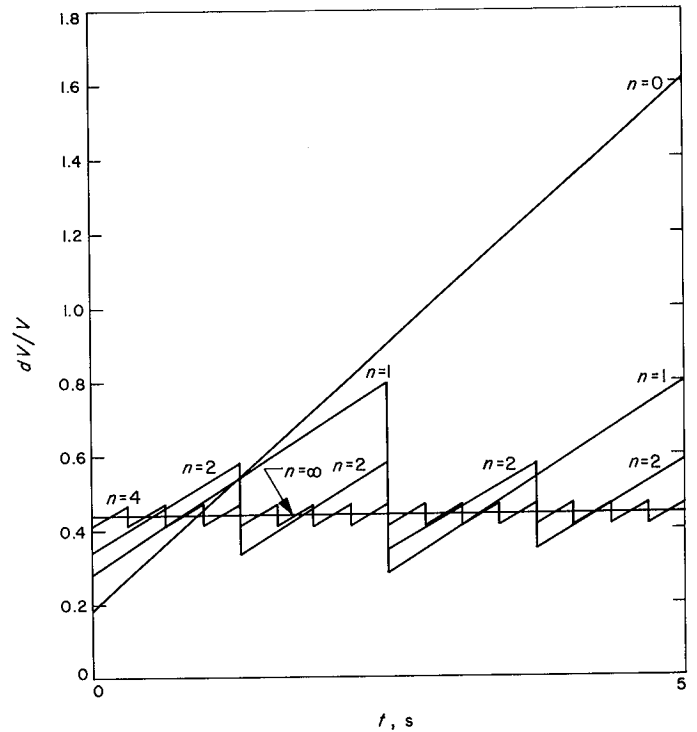


Fig. 2. The ratio of slope to voltage as a function of n

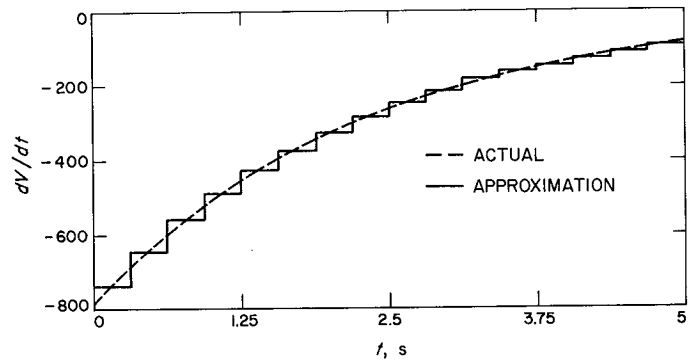


Fig. 3. The rate of change of voltage for the exponential and the approximation

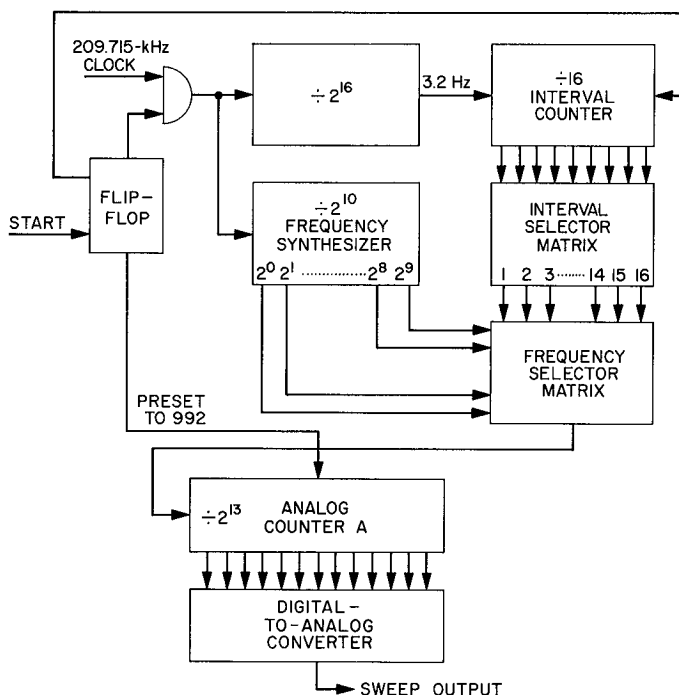


Fig. 4. Block diagram of sweep generator

The method selected to generate the approximation curve is a binary counter linked to a digital-to-analog converter with the counter driven by a frequency synthesizer. The block diagram is shown in Fig. 4.

Table 4 is a computer printout of the end points of the line segments. The computed frequency is based on the assumption of a 0.25-V output change per count. The high-

Table 4. End points of line segments

Interval	Voltage, V		f, Hz
	P ₁	P ₂	
1	1800.0000	1569.0341	2956.3635
2	1569.0341	1367.7042	2577.0227
3	1367.7042	1192.2080	2246.3513
4	1192.2080	1039.2304	1958.1132
5	1039.2304	905.8822	1706.8569
6	905.8822	789.6443	1487.8440
7	789.6443	688.3216	1296.9314
8	688.3216	600.0000	1130.5167
9	600.0000	523.0113	985.4547
10	523.0113	455.9014	859.0073
11	455.9014	397.4026	748.7837
12	397.4026	346.4101	652.7040
13	346.4101	301.9607	568.9529
14	301.9607	263.2148	495.9477
15	263.2148	229.4405	432.3107
16	229.4405	199.9999	376.8387

est frequency required is 2956 Hz and the next highest is 2577 Hz. The clock frequency should be such that the incremental sweep interval, 0.3125 s, can be obtained by dividing the clock by a power of 2. Thus

$$\frac{F_{clock}}{2^k} = \frac{1}{0.3125 \text{ s}} = 3.2 \text{ Hz}$$

$$= 3.2 \cdot 2^k$$

and

$$2956 = \frac{F_{clock}}{n_1}$$

$$2577 = \frac{F_{clock}}{n_2}$$

where n_1 and n_2 are integer values dividing the clock to produce the two frequencies

$$n_1 = \frac{3.2 \cdot 2^k}{2956}$$

$$n_2 = \frac{3.2 \cdot 2^k}{2577}$$

Obviously n_1 must be smaller than n_2 , and the difference between them must be an integer:

$$n_2 - n_1 \geq 1, n_2 - n_1$$

$$= 2^k \left(\frac{2956 - 2577}{2956 \cdot 2577} \right) 3.2$$

$$= 3.2 \cdot 2^k \left(\frac{379}{7.54 \cdot 10^6} \right) = \frac{1210 \cdot 2^k}{7.54 \cdot 10^6}$$

$$= \frac{1.21 \cdot 10^3 \cdot 1.024 \cdot 10^3 \cdot 2^{k-10}}{7.54 \cdot 10^6}$$

$$= \frac{1.24 \cdot 2^{k-10}}{7.54}$$

$$= \frac{9.1 \cdot 2^{k-13}}{7.54} = 1.3 \cdot 2^{k-13}$$

Therefore, the minimum value of k is 13, and $2^{13} = 8192$. If k is allowed to be 13 and $n_1 - n_2$ is an integer, 1.3 must be rounded off to 1, a 30% error. By increasing k , the round-off error can be reduced as shown in Table 5. For the breadboard, the value chosen for k was 16. Thus

$$F_{clock} = 3.2 \cdot 2^{16} = 209,715.2 \text{ Hz}$$

Table 5. Reduction of round-off error

k	$n_2 - n_1$	Round-off error, %
13	1.3	30
14	2.6	13.3
15	5.2	4
16	10.4	4
17	20.8	1

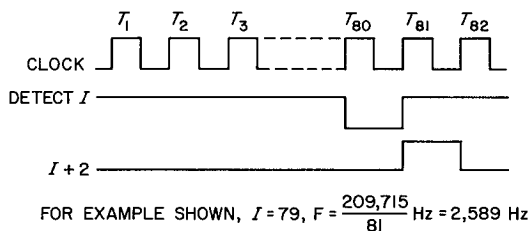
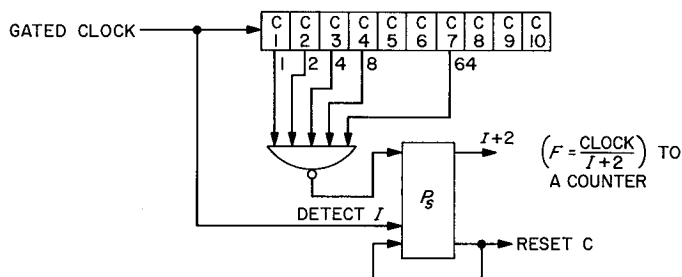


Fig. 5. Frequency synthesizer operation

The basic operation of the synthesizer is shown in Fig. 5. The division is done by the C counter, which counts until it reaches a number I that is two less than the divisor n . When the counter reaches I , the next clock pulse sets the flip-flop pulse P_s which, in turn, resets the C counter. The next clock pulse later, P_s is reset, and the C counter starts again. The pulse is therefore occurring every $I + 2$ clock pulses, which is equal to the clock divided by n . Table 6 lists the calculated n , the integer value I , the effective value of n , the number of P_s pulses that would result from using the calculated value of n , and the actual number of P_s pulses.

The T counter divides the clock by 2^{16} to give an output of 3.2 Hz to drive the interval counter. Sixteen gates decode the C counter and the interval counter simultaneously to give the proper I output during the appropriate interval. The digital-to-analog converter reference can be set to produce a sweep from 18 to 2 V or 9 to 1 V. Multi-

Table 6. Calculated and actual values of P_s

Interval	n		I	P_s	
	Calculated	Effective		Calculated	Actual
1	70.92	71	69	924	922
2	81.34	81	79	805	809
3	93.33	93	91	702	704
4	107.1	107	105	612	612
5	122.9	123	121	533	535
6	140.9	141	139	465	468
7	161.7	162	160	405	404
8	185.5	186	184	353	352
9	212.8	213	211	308	307
10	244.0	244	242	268	269
11	279.9	280	278	234	234
12	321.2	322	320	204	203
13	369.4	370	368	178	177
14	422.6	422	420	155	155
15	484.9	485	483	135	135
16	556.4	556	554	118	118

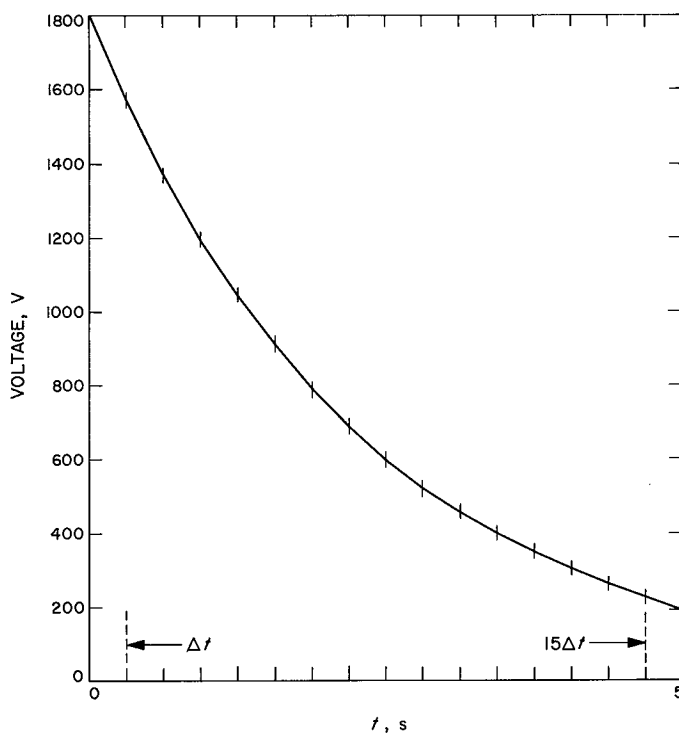


Fig. 6. The approximated sweep voltage

plication of this output by 100 or 200 will give the output shown in Fig. 6, the approximation of $1800 \exp(-0.44t)$.

PRECEDING PAGE BLANK NOT FILMED.

XVI. Lunar and Planetary Sciences

SPACE SCIENCES DIVISION

A. Solar Wind Interaction With Solids, *H. C. Lord*

Low energy gas ions incident upon solids will form a gas-rich surface layer on these objects. The solar wind provides a source of low energy ions. At 1 AU the solar wind flux is about 2×10^8 particles/cm² s, and the energy of the proton component (the most abundant species) has a mean value slightly above 1 keV (Ref. 1). Thus, meteorites, cosmic dust, and, most likely, the lunar surface are irradiated in this manner.

To further understand this phenomenon, silicate samples were irradiated in this work, with 2-keV protons and 1.8-keV helium ions in separate experiments. These irradiated samples were then degassed, with quantitative measurement of the concentrations of the retained gases using a gas chromatograph. Knowing the incident dose, retention coefficients were calculated, and by plotting retention coefficients versus incident dose the saturation value for the injected gas in the substrate was determined. A step-wise heating procedure was used, with the amount of released gas measured at each temperature. In this way a gas release curve was obtained, which is characteristic of the incident ion species and dose, as well as the surface temperature of the substrate.

It was found that substantial hydrogen and helium are retained by olivine and enstatite under these conditions, as shown in Fig. 1.¹ The saturation value for 2-keV protons in forsterite is about 5×10^{17} cm⁻², and for 1.8-keV helium ions in forsterite is about 6×10^{16} cm⁻². These imbedded atoms are in a layer less than 500 Å thick (Ref. 2). Assuming a density of 3.5 g/cm³ and a mean atomic weight of 20 for these silicates, this layer contains 5×10^{17} lattice atoms/cm² of top surface area. Surface roughness may increase this value somewhat; but, nevertheless, the trapped gas atoms are present at saturation in an approximate 1 to 1 correlation with the lattice atoms.

Maximum gas release (in cm³/cm² irradiated area/100°C temperature interval) of these low energy injected ions occurs at or below 400°C. The temperature profiles for the release of 2-keV protons injected into forsterite, and the results for the release of 1.8-keV injected helium ions in forsterite are shown in Fig. 2.

¹The approximate value of the retention coefficient for each irradiation can be read. The higher flux irradiations show decreased concentration of desorbed gas with increasing incident dose due to increasing sample surface temperature.

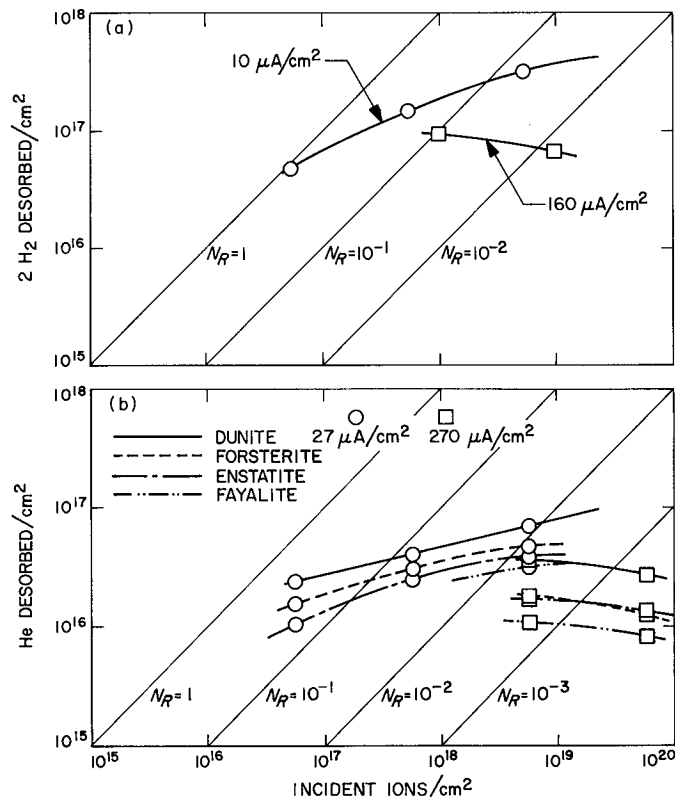


Fig. 1. Gas release curves: (a) desorbed hydrogen from forsterite as a function of the incident dose of 2-keV protons; (b) desorbed helium as a function of the incident dose of 1.8-keV He^+

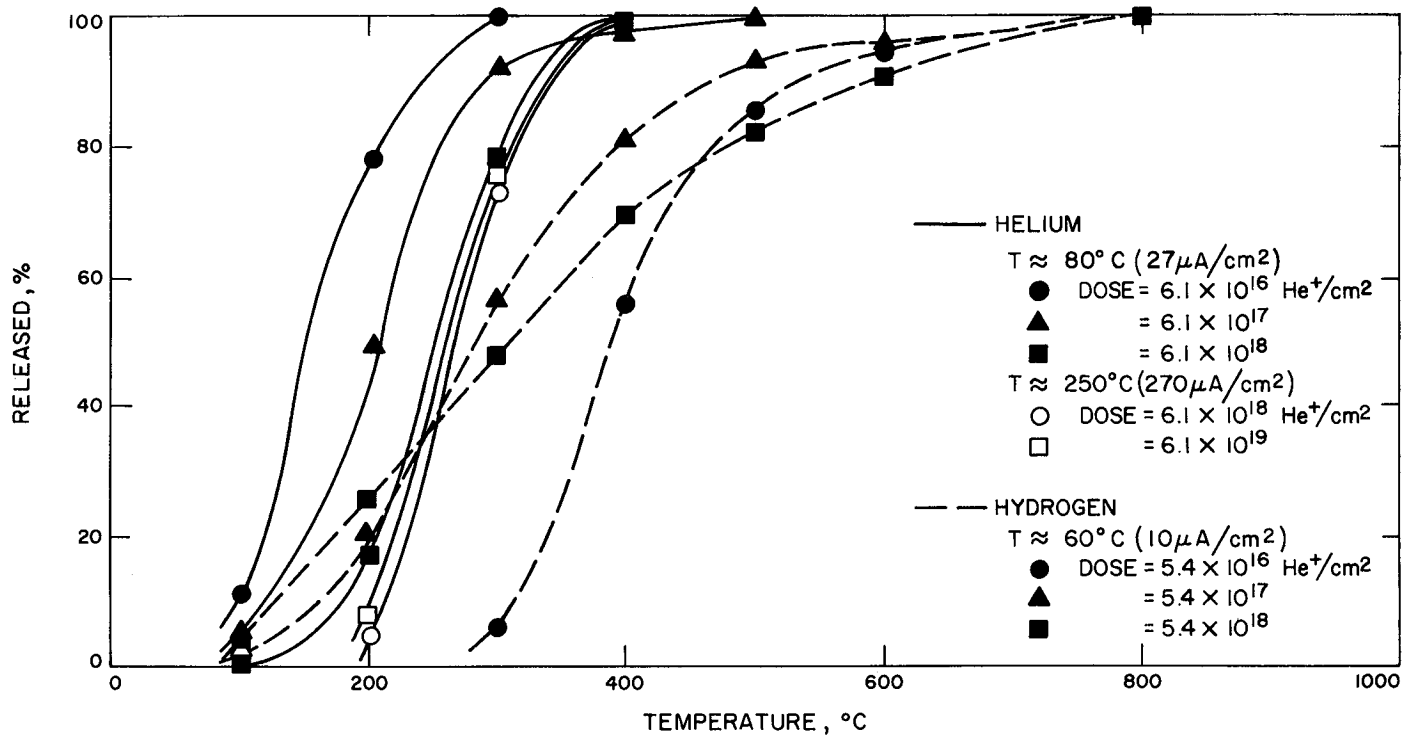


Fig. 2. Integral gas release curves for a range of doses of 1.8-keV helium ions and 2.0-keV hydrogen ions in forsterite

The higher release temperature for hydrogen as compared to helium for the conditions used here, as well as the preponderance of hydrogen in the solar wind, indicates that solar wind injected hydrogen should be present in those samples showing solar wind trapped helium (such as gas-rich meteorites). If the low temperature released contamination hydrogen can be separated out, a method is provided to obtain information about both the solar wind hydrogen to helium ratio and also the solar wind hydrogen to deuterium ratio at the time when the gases were implanted. For a quantitative analysis of the results of this postulated experiment, it will first be necessary to determine the relative retention coefficients of hydrogen, deuterium, and helium in the presence of each other. These experiments will be performed in the near future.

The detection and measurement of solar wind injected gas in a solid indicates an extra terrestrial origin for the sample. The contamination, radiogenic, and cosmogenic gas components, if present, must be separated from the total released gas. This can be done, as with the gas-rich meteorites, by consideration of the total gas concentration, isotopic composition of the released gas, and, if necessary, the release curves of the various components of the gas. For cosmic dust of grain size about $10\text{ }\mu\text{m}$, each grain if saturated with solar wind injected helium would contain about $8 \times 10^{-9}\text{ cm}^3$. The molecular hydrogen concentration would be five times higher.

A more detailed account of this work has been submitted for publication.

References

1. Neugebauer, M., and Snyder, C. W., "Mariner II Observations of the Solar Wind: 1. Average Properties," *J. Geophys. Res.*, Vol. 71, pp. 4469-4484, 1966.
2. Hines, R. L., and Arndt, R., "Radiation Effects of Bombardment of Quartz and Vitreous Silica by 7.5-keV to 59-keV Positive Ions," *Phys. Rev.*, Vol. 119, pp. 623-633, 1960.

B. 1967 Radar Observation of Mars, R. L. Carpenter

As part of the continuous DSN research program, Mars was observed by radar throughout almost every night of April and May, 1967.² Radar parameters were as follows:

Radiated power: 100 kW
Two-way antenna gain: 108.5 dB
Wavelength: 12.5 cm
System noise temperature: 20°K

Occasionally, the 210-ft antenna at Mars deep space station became available for reception. This provided an improvement of 7.5 dB in SNR, which resulted in greatly improved data.

The collected data were all in the form of spectrograms. Monochromatic waves were beamed at Mars, and the doppler-broadened spectra of the echos were recorded. Altogether, 1088 spectra were received with the 85-ft antenna and 141 with the 210-ft. In this article, only the bistatic (210-ft antenna) runs are analyzed. These spectra have been combined into approximately 60 composite spectra representing the average spectrum of Mars for every 5 deg of longitude of the planet. The spread in longitude for each composite spectrum is 10 deg.

Figure 3 shows three examples of the spectra that were obtained. Figure 3(a) is typical in appearance while (b) is

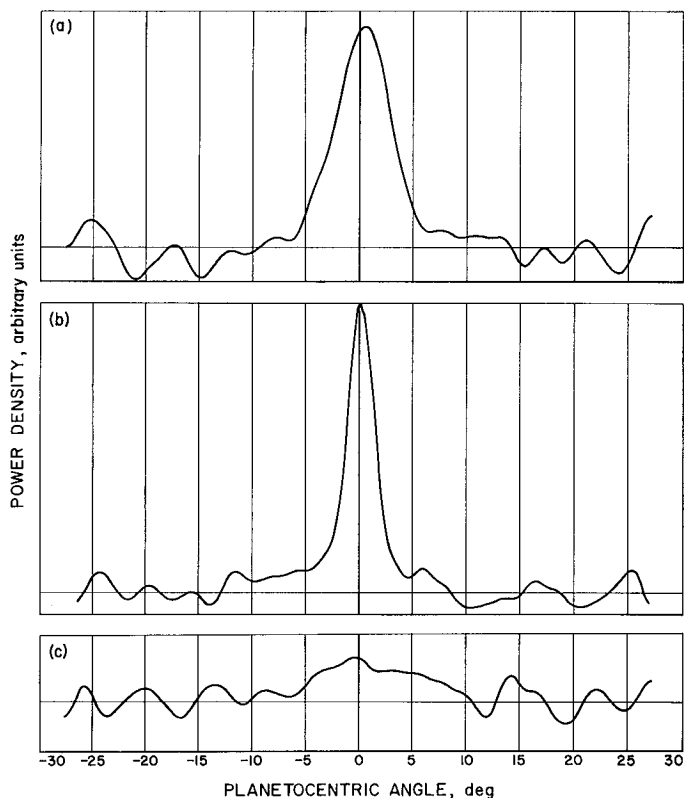


Fig. 3. Typical radar spectra of Mars: (a) longitude 4.5 deg; (b) longitude 199.3 deg; (c) longitude 81.7 deg

²The experiments were performed by the author (then of JPL Section 331, now of Section 325) and R. M. Goldstein (Section 331) with the help of E. Jackson's Goldstone Support Group (Section 335).

the narrowest spectrum obtained and also has the strongest central peak. Figure 3(c) is one of the weakest spectra and shows little or no central peak. The ordinate is power density in arbitrary units; however, all three spectra have the same scale. The abscissa is expressed in planetocentric degrees centered on the subearth point and is measured along the doppler equator of Mars.

The maximum bandwidth that could be employed during the experiment was 3700 Hz. Since the doppler bandwidth of Mars is about 7130 Hz, only the central 52% of the total spectrum was observed, which corresponds to $\pm \sim 30$ deg in longitude. This restriction in coverage is of little consequence since the reflected power from regions at ± 30 deg and beyond is so weak as to be lost in the noise.

The base line in each spectrum was found by fitting a horizontal line to those regions beyond $\pm \sim 22$ deg. Again, because the reflected power is so weak at the edge of the spectra, the fitted base line and the true zero of the spectra are practically coincident. More than 90% of the total reflected power is obtained with this base line fitting procedure. This does mean, however, that the radar cross-sections given below may be low by approximately 10%.

The half-width of the typical spectrum (Fig. 3(a)) at its half power point is about 3 deg. This can be taken as a rough approximation to the median slope of the martian surface, averaged over all longitudes for the region scanned by the radar and for areas larger than a few wavelengths across. The latitude scanned was about +21 deg. Spectrum (b) corresponds to the region near Trivium Charontis. Its half-width is only $1\frac{1}{2}$ deg. The half-width of (c) is difficult to assess but appears to be greater than 7 deg. It is apparent that there are large variations in the roughness of the topography of Mars.

Figure 4 shows several characteristics derived from the spectra and their relation to the visual appearance of Mars. At the bottom of Fig. 4 is a drawing of the visual feature on Mars for the region scanned during the experiment. The drawing was made from the map prepared by Dollfus (Ref. 1). The heavy black line shows the course of the subearth point as Mars rotated. Figure 4(a) shows the relative radar cross section as a function of longitude. The relative radar cross-section is the ratio of the observed reflected power to that which would be expected from a smooth sphere the same size as Mars with a surface reflectivity of unity. It is derived from the known radar system parameters, the range to Mars,

and the area of the spectra. In Fig. 4(b) is shown the height of the peak of the spectra in arbitrary units. Since the height of the peak is determined primarily by the reflectivity of the martian surface at the subearth point, it is a more accurate measure of the way the reflectivity changes with longitude. A measure of the smoothness or specularity of the surface can be obtained from the half-width of the spectrum at the half power point. Figure 4(c) shows the reciprocal of the half-width expressed in planetocentric degrees. As mentioned above, it can be considered as an estimate of the median slope of the surface in the region of the subearth point. The gap in the plots between 105 and 170 deg in longitude is due to the fortuitous lack of bistatic observations when this region was crossing the center of the martian disk.

The mean radar cross-section is 6.3%, which is close to that found for the moon, about 7%. A remarkable feature of the radar cross-section plot is its variability as a function of longitude. It ranges between 1.5 and 12.3%. By comparison, the radar cross-section of Venus has been observed over a planetocentric longitude of approximately 90 deg, and it does not vary more than a few percentage points about its mean of 11.2%. This suggests that the equatorial region of Mars is significantly more heterogeneous than that of Venus. The very low radar cross-sections are particularly puzzling. For example, if the dielectric constant of the minerals on Mars is approximately 8, which is typical of terrestrial minerals, then to obtain a radar cross-section of under 2% requires that the minerals be pulverized to such an extent that the porosity of the material is greater than 80%. About the only terrestrial substance with this high a porosity is silt. The largest reflectivity is found in the region of Laocoöntis (245 deg), which is visually relatively inconspicuous. The smallest reflectivities are found from 70 to 105 deg in longitude and perhaps beyond. This is in a light region showing a few vague features. Note also that this region is the roughest, as indicated by the half-width plot, Fig. 4(c).

The subearth point was scanned through four dark regions: Trivium Charontis (195 deg), Nepenthes (260 deg), Syrtis Major (285 deg), and the southern part of Mare Acidalius (40 deg). Curiously, there are local minima in the cross section and height of the central peak for the latter three features. Also, the halfwidths appear to show local maxima as well. This could suggest that the dark regions are intrinsically poorer reflectors and somewhat rougher than the surrounding lighter regions. This does not imply, however, that the light regions are generally better reflectors and smoother. For

example, the major light regions of Elysium and Tractus Albus have very low radar cross-sections and are quite rough. Note also the extreme difference in the radar characteristics of the two light regions Elysium (western side, 215 deg) and Isidis Regio (270 deg). As noted previously, the smoothest region is near Trivium Charontis; however,

the minimum bandwidth is not coincident with it. The minimum bandwidth is more likely associated with an area called Albor (200 deg). Albor is a relatively small, very light region at the east end of Elysium. It is not shown on the map of Dollfus but has been seen many times by other visual observers (Ref. 2).

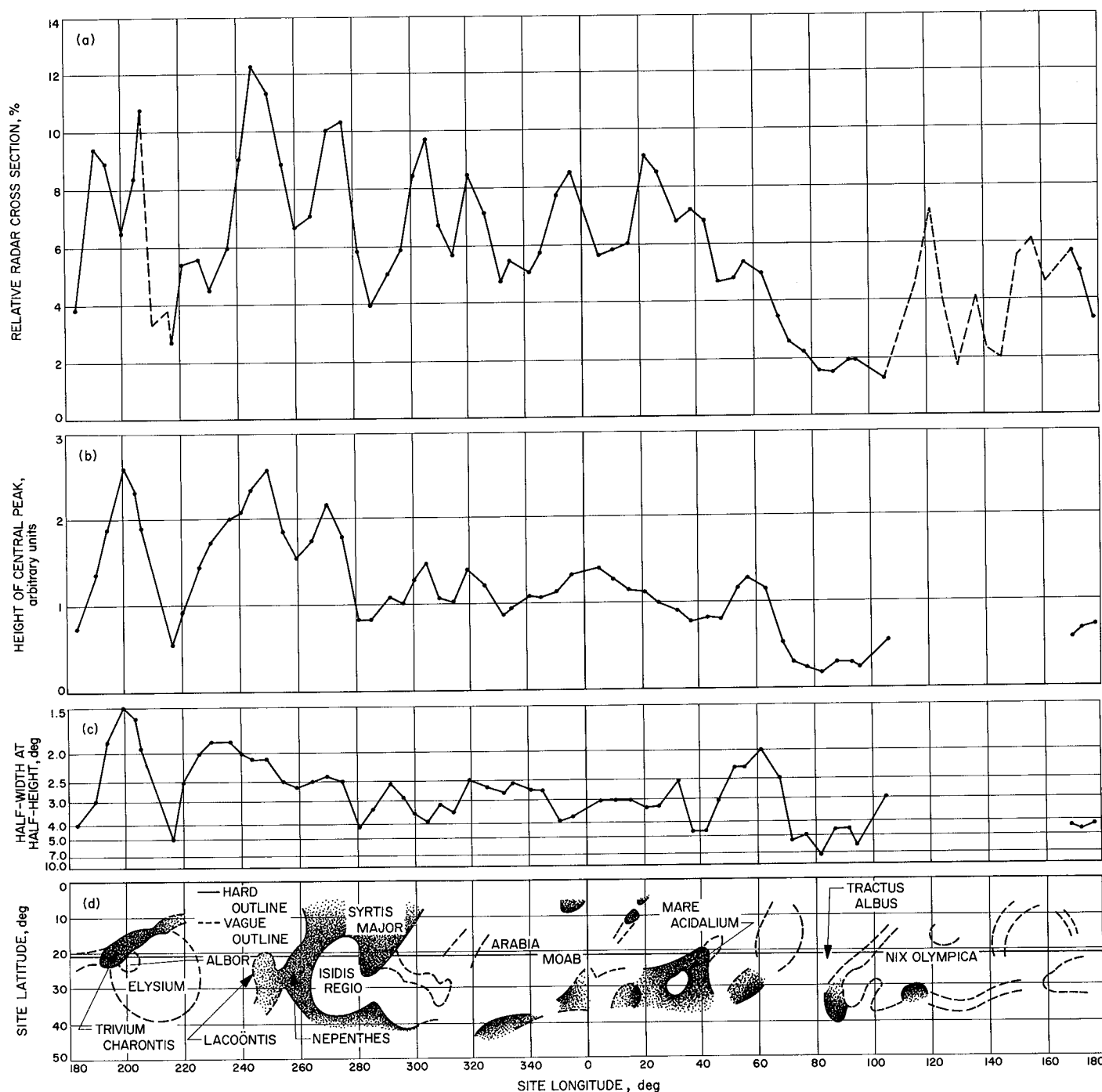


Fig. 4. Relation between radar measurements and visual features on Mars: (a) relative radar cross-section; (b) height of central peak; (c) half-width at half-height; (d) site latitude vs longitude

It would appear that the visual and radar appearance of Mars show no clear relationship and that each corresponds to a different aspect of the planet's surface.

References

1. *Planets and Satellites*. Edited by G. P. Kuiper, University of Chicago Press, 1961.
2. Capen, C. F., *The Mars 1964-1965 Apparition*, Technical Report 32-990, Jet Propulsion Laboratory, Pasadena, Calif., Dec. 15, 1966.

C. Possibility of Permafrost Features on the Martian Surface, F. A. Wade³ and J. N. de Wys

Permanently frozen ground is a distinct possibility at all latitudes on Mars. As shown in Fig. 5, 10- to 20-deg

³Professor of Geology, Texas Technological College, Lubbock, Texas.

latitudes sustain the longest diurnal rise periods, with temperatures rising daily above 0°C. Poleward from

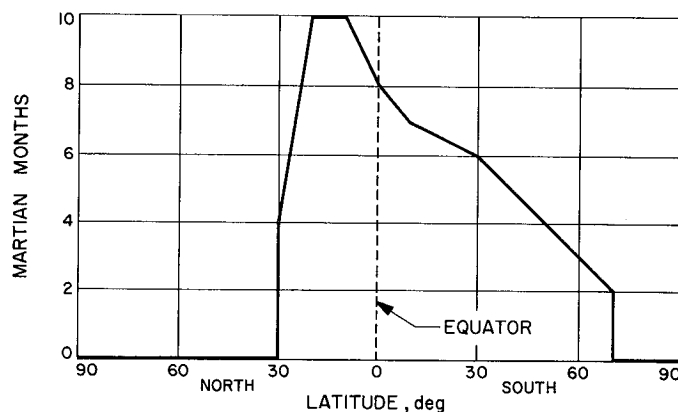


Fig. 5. Number of Martian months during which temperatures rise diurnally above 0°C vs latitude

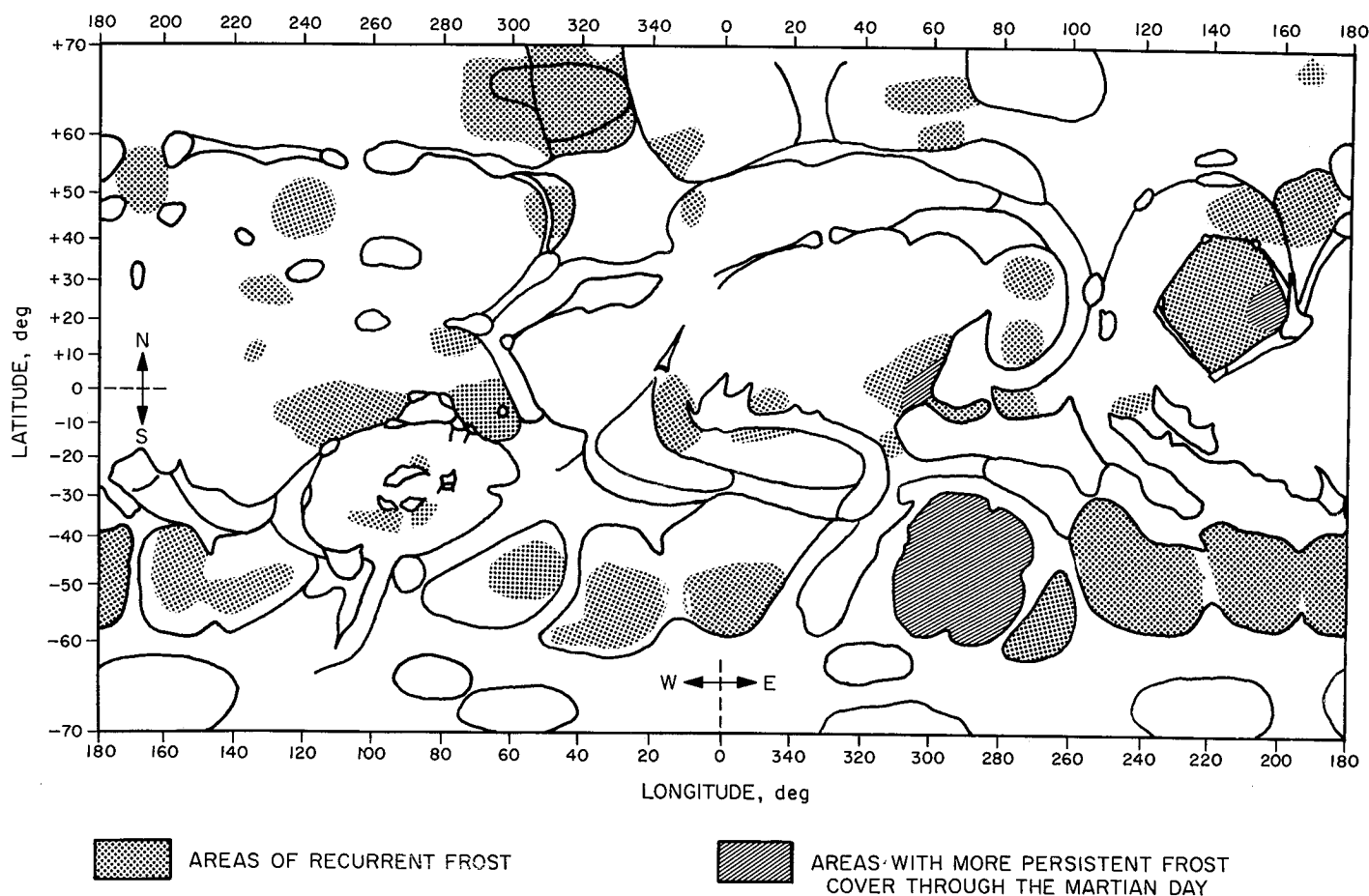


Fig. 6. Outline map of Mars showing areas of recurrent frost

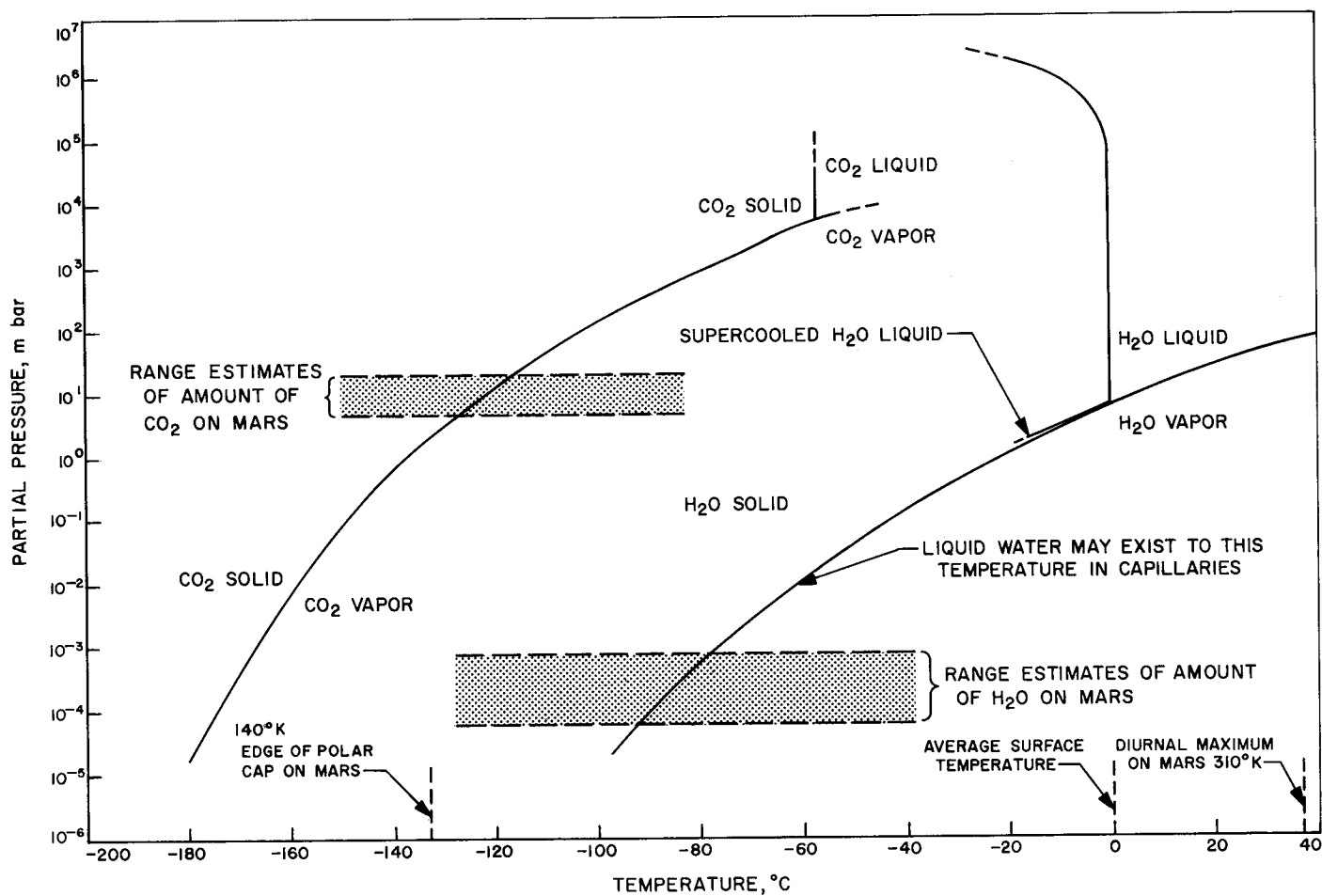


Fig. 7. Phase relationships of carbon dioxide and water

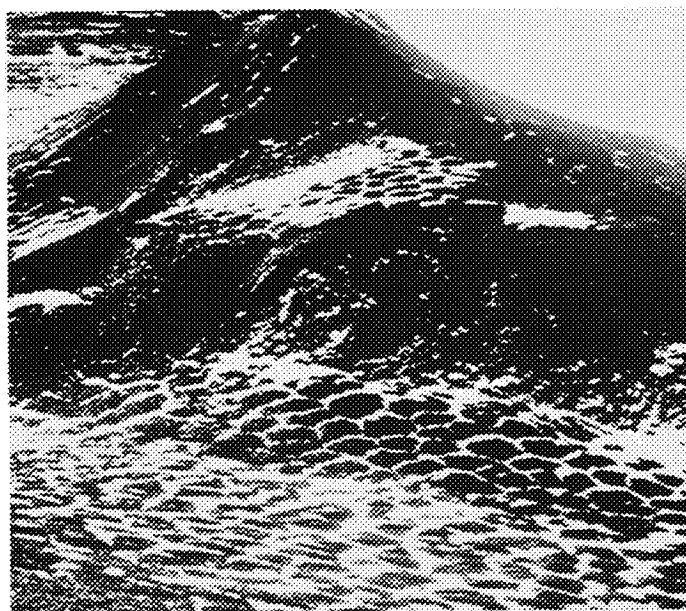


Fig. 8. Nonorthogonal patterned ground developed in ground moraine, Shenk Peak area, Queen Maud Range, Antarctica

30 deg N latitude and 70 deg S latitude, however, the temperature is always below 0°C.⁴ The mean shallow subsurface temperatures are below 0°C.

The areas of apparent recurrent frost on Mars are shown in Fig. 6. Not all areas show apparent frost simultaneously. The fact that some white areas remain throughout the martian day and pass through the subsolar point suggests that the composition of the frost is water rather than carbon dioxide.⁵ This is further suggested by Fig. 7, which shows the phase relationships of carbon dioxide and water with range estimates of amounts present on Mars indicated by stippled areas. It can be seen that, at tempera-

tures higher than about -120°C, the solid phase (frost) must be composed of water.

Because of the probable absence of liquid water on the surface, any ice-saturated permafrost would have formed from water reaching the surface zone from the interior during volcanism or outgassing. If the frost is composed of water, either meteoric or juvenile, ridges and troughs arranged in polygonal patterns and irregular mounds (formed by frost heaving) may be present in the permafrost areas. An example of the type of polygonal-patterned ground is shown in Fig. 8. For the type of terrain shown, the polygons may be up to 30 m in diameter and 1.5 m deep, with rims up to 2 m high and 1.5 m wide. Similar developments on Mars may be caused by sand wedges with no ice, because of the seasonal expansion and contraction of the surface with alternate penetrations of heat and cold waves. Any water or hoar frost crystals forming in fractures would aid this process.

⁴Leighton, R. B., Theoretical Computer Model of Martian Surface Temperatures, private communication, 1967.

⁵As observed by C. Capen.

XVII. Bioscience

SPACE SCIENCES DIVISION

A. Picric Acid Stability in Aqueous Sodium Hydroxide as Related to the Biosatellite

Mission, J. P. Hardy and J. H. Rho

1. Introduction

An aqueous alkaline solution of picric acid is used as the color forming reagent in biological assays of creatinine in urine (Ref. 1). Under normal laboratory conditions, aqueous solutions of picric acid and sodium hydroxide are mixed, in the course of an analysis, to make the color-forming reagent. However, to meter and mix reagents in an automated instrumental assembly is no simple task. To obviate an unnecessary mixing step in the creatinine analysis, the reagents, sodium hydroxide and picric acid, can be pre-mixed. The question then arises as to the stability of this mixed *picrate* reagent.

2. Literature

V. Gold and C. H. Rochester (Ref. 2) have studied the reactions of very dilute solutions of picric acid (10^{-5} M) in sodium hydroxide at 40°C. Since picric acid is a strong acid [$pK_a = +0.38$ at 25°C (Ref. 3)], even in water solution it exists entirely as picrate ion. This picrate

anion has an absorption maximum at 3600 Å. In more concentrated sodium hydroxide solutions, above 0.4 M, a new picrate species is formed with an absorption maximum at 3900 Å. At 3.65 M sodium hydroxide, the picrate is converted completely to the new species.

Gold and Rochester report that when visible light is excluded from a reaction vessel, the intensity of each of the absorption maxima decreases with time. The calculated first order rate constant (first order with respect to picrate) for the reaction is dependent upon hydroxide concentration; the rate constant is larger at higher concentrations (Table 1). The decrease in absorption intensity is accompanied by the production of nitrite ions;

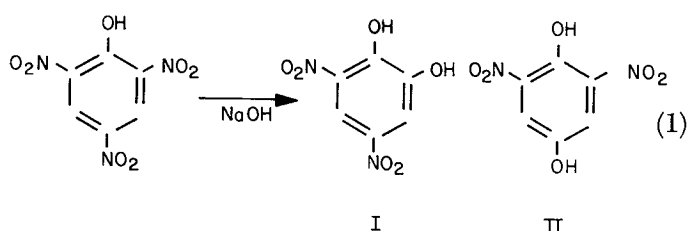
Table 1. Rate constants for decomposition of 3.91×10^{-5} M picric acid in aqueous sodium hydroxide solutions at 40°C^a

[NaOH], M	0.394	0.789	1.182	2.65	3.44
$10^7 k$, s ⁻¹	3.07	6.12	11.7	35.4	37.5

^aData taken from Ref. 2.

approximately one nitrite is produced per molecule of picric acid initially present in the solution. In intense visible light, the rate of decrease of the absorption maximum and the rate of production of nitrite ion are increased.

The data suggest that some form of picrate is undergoing a nucleophilic substitution reaction by hydroxide to produce dihydroxydinitrobenzenes and nitrite. Reasonable structures for the substitution products are 3,5-dinitrocathechol (I) and 2,6-dinitroquinol (II), Eq. (1).



In basic solution, 2,6-dinitroquinol is unstable and is converted to a species with an absorption maximum at 3900 Å. The rate of conversion is increased by the presence of oxygen which suggests that 2,6-dinitroquinol is being oxidized by 2,6-dinitroquinone. Under the same conditions, 3,5-dinitrocathechol is stable.

3. Results and Discussion

a. Alkaline picrate stability. That alkaline picrate is unstable is established; the questions of importance which relate to an automated analytical procedure are questions of degree. At the ambient temperatures encountered in an extended biosatellite flight, to what degree will the picrate reagent decompose? Will the degradation products interfere with the analysis?

As was indicated, the concentration of picrate used by Gold and Rochester in their extensive studies was low, i.e., 10^{-5} M. On the other hand, the picrate reagent, is made from a saturated aqueous solution of picric acid which is diluted four-fold with sodium hydroxide to give a solution containing 0.0145 M picrate and 0.575 M base.

To determine the effect of picrate concentration on the reaction kinetics, and, also, to study the effect of various other factors such as light and the presence of foreign materials,¹ a kinetic study of alkaline picrate decomposition was made at $74.00 \pm 0.05^\circ\text{C}$. Individual samples were sealed in glass ampoules which were immersed in

a constant-temperature water bath. The reaction was monitored by measuring the disappearance of the absorption maximum at 3600 Å. In all cases, the entire spectrum of each sample between 7000–2500 Å was recorded.

As anticipated, the picrate concentration showed an exponential decay; the first order rate constant at 75.0°C was $1.86 \pm 0.36 \times 10^{-5} \text{ s}^{-1}$, which corresponds to a 0.43-day half-life. Within experimental error for any given point, $\pm 2.5\%$, all extraneous factors had no effect on the rate. Tubes wrapped in aluminum foil or left exposed to room light contained the same concentration of picrate. If the picrate solution was made 0.0285 M in sodium chloride, no change in the rate was seen. The inclusion of silastic rubber tubing or uncured silastic rubber sheet showed no effect. The lack of effect of these latter materials suggests that the picrate reagent is compatible with the materials intended for flight packaging.

That light has very little effect in the present situation whereas Gold and Rochester reported the reaction was catalyzed by light is most likely a reflection of the short reaction times at 75°C , and the difference in the level of light intensities. Gold and Rochester irradiated a 2-cm path-length quartz cell with a 150-W tungsten filament source. In contrast, the present work was carried out with 2-ml aliquots of the much more highly concentrated solution sealed in borosilicate tubes which were immersed beneath at least 10 cm of water, and illuminated by normal room lighting.

Once the general characteristics of the picrate degradation reaction had been determined, long term studies were carried out at temperatures expected to prevail during the biosatellite flight.

Over a period of 14 days at $30 \pm 1^\circ\text{C}$, the alkaline picrate reagent decomposed less than 15% as measured by the decrease in the 3600 Å absorption maximum. The calculated first order rate constant is $8.6 \pm 3.3 \times 10^{-8} \text{ s}^{-1}$, which corresponds to a half-life of 93 days.

At 41°C , the maximum temperature anticipated during a biosatellite mission, the picrate reagent decomposes only 22% in 6.7 days.² A first order rate constant is thus approximately $3 \times 10^{-7} \text{ s}^{-1}$. This corresponds to a half-life of 27 days.

b. Creatinine assay with alkaline picrate. Although it is convenient to measure the picrate concentration by

¹Silastic rubber, sodium chloride, etc.

²Measured using the 3600 Å absorption maximum.

observing the intensity of the 3600 Å maximum, such a measurement gives no indication whether the partially degraded picrate is still usable as a color forming reagent for creatinine analysis. Indeed, interference from the degradation products might be expected since these products are structurally similar to picric acid. A second problem, in this respect, concerns the stoichiometry and equilibrium constant for the formation of the picrate-creatinine complex, Eq. (2).

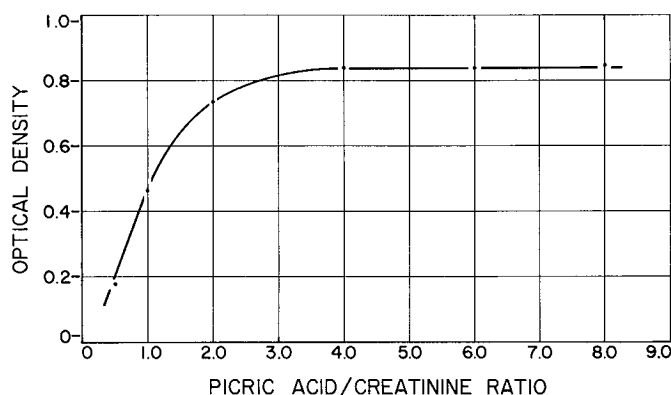
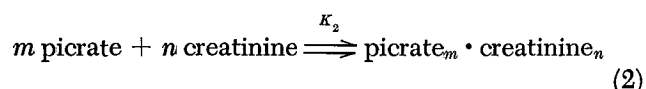


Fig. 1. Plot of optical density at 4800 Å against picric acid to creatinine ratio for 10^{-3} M creatinine

Addressing the second problem first, a series of solutions containing a constant amount of creatinine, 10^{-3} M, was analyzed in the usual way, except that the picrate concentration was varied; the picrate to creatinine ratio extended from 0.5 to 8.0 (Fig. 1). An examination of Fig. 1 shows that at least a four-fold excess of picrate reagent is needed for the analysis. On the basis of the data in Table 2 and Fig. 1, and assuming the colored complex is a one to one complex of picrate and creatinine ($m = n = 1$, Eq. 2), a rough value for K_2 is calculated to be $3.2 \pm 1.8 \times 10^3 \text{ M}^{-1}$.

To determine the effect of picrate degradation products on the colorimetric assay, a creatinine solution, 10^{-3} M, was analyzed with various samples of partially degraded picrate reagent. These values were compared with the results obtained using fresh picrate solutions. The results are recorded in Table 3. As expected, partial degradation of the picrate reagent, up to 22% loss, has no effect on the creatinine assay. However, after a substantial amount of degradation, more than 46% loss, the assay for creatinine drops off rapidly. If the degradation products did

Table 2. Effect of picric acid to creatinine concentration ratio on the creatinine assay^a

Creatinine, M	Picrate, M	Picrate creatinine	Absorbance (4800 Å)
10^{-3}	0.5×10^{-3}	0.5	0.177
10^{-3}	1.0×10^{-3}	1.0	0.460
10^{-3}	2.0×10^{-3}	2.0	0.730
10^{-3}	4.0×10^{-3}	4.0	0.835
10^{-3}	6.0×10^{-3}	6.0	0.833
10^{-3}	8.0×10^{-3}	8.0	0.845

^aAssay solution contains 0.2 N NaOH, 0.04 N NaCl, and 0.4 N Na₂SO₄.

Table 3. Rates of degradation of picric acid in aqueous sodium hydroxide solution

Temperature, °C	Container	Time, days	Decomposition, %	% of maximum creatinine analysis
30 ± 1	Glass ampoule	13.95	15	98.6
41 ± 0.1	Silastic bag	0	0	99.2
41	Silastic bag	1.4	13	97.9
41	Silastic bag	3.7	40–50 ^a	29.2
41	Silastic bag	5.62	— ^a	0.5
41	Glass ampoule	0	0	99.8
41	Glass ampoule	2.6	2	100.4
41	Glass ampoule	4.67	11	102.2
41	Glass ampoule	5.85	17.9	100.4
41	Glass ampoule	6.71	21.7	103.4
41	Glass ampoule	9.73	46.0	85.9
41	Glass ampoule	10.77	48.5	87.1
41	Glass ampoule	10.94	—	79.3
41	Glass ampoule	11.74	54.2	73.3
41	Glass ampoule	16.76	69.0	2.6

^aAbsorption maximum shifted from 3600 to 3900 Å.

not interfere with the analysis at all, one would expect that the reagent could degrade by 73% before any effect was seen on the creatinine analysis.³ The degraded picric acid solution is dark-brown in color which is a reflection of the absorption of the degradation products in the visible region. The presence of this dark material inevitably interferes with the creatinine analysis to some extent, for the absorption of this species tails well into the 4800 Å region where the picrate-creatinine complex has its absorption maximum. This, of course, increases the absorbance of the blank (picrate reagent without creatinine) and introduces an error, since the result of the determination is read as the difference between two large numbers.

However, the extent to which the creatinine assay decreases as the picrate reagent degrades is greater than can be accounted for by a shift in the equilibrium (Eq. 2) or the error introduced by the increased blank reading.

4. Conclusions

The rate of decomposition of picric acid in aqueous sodium hydroxide was studied at 75°, 41°, and 30°C to determine the characteristics of the decomposition reaction. At 75°C, the reaction was observed to follow first-order kinetics up to 82.6% reaction. The calculated first-order rate constant was $1.86 \pm 0.36 \times 10^{-5} \text{ s}^{-1}$.

At 41°C, the maximum temperature anticipated during the biosatellite mission, the picrate reagent decomposes

³A 73% decrease in the concentration of 0.0145 M picrate gives 0.004 M picrate, which is still the requisite four-fold excess over the 0.001 M creatinine being assayed.

only 22% in 6.7 days as measured using the 3600 Å absorption maximum. The first-order rate constant is thus approximately $3 \times 10^{-7} \text{ s}^{-1}$. This corresponds to a half-life of 27 days.

Over a period of 14 days at $30 \pm 1^\circ\text{C}$, the alkaline picrate reagent decomposed less than 15%. The calculated first-order rate constant is $8.6 \pm 3.3 \times 10^{-8} \text{ s}^{-1}$ which corresponds to a half-life of 93 days.

To determine the effect of picrate degradation on the colorimetric assay, a creatinine solution was analyzed with various samples of partially degraded picrate reagent. This study showed that the alkaline picrate reagent is sufficiently unstable at 41°C that its decomposition interferes with the creatinine analysis after 7 or 8 days. The extent to which the creatinine assay decreases as the picrate reagent degrades is greater than can be accounted for by the loss of picrate. These findings prompted a change in the biosatellite automated assay; the picrate reagent is freshly mixed just before analysis.

References

1. Taussky, H. H., "Creatinine and Creatine in Urine and Serum," *Standard Methods of Clinical Chemistry*, Vol. 3, pp. 99-113. Edited by D. Seligson, Academic Press, New York, 1961.
2. Gold, V., and Rochester, C. H., "Reactions of Aromatic Nitrocompounds in Alkaline Media. Part VII. Behaviour of Picric Acid and of Two Dihydroxydinitrobenzenes in Aqueous Sodium Hydroxide," *J. Chem. Soc.*, pp. 1722-1727, 1964.
3. Weast, R. C., ed., *Handbook of Chemistry and Physics*, 47th Edition, p. D-87, The Chemical Rubber Co., Cleveland, Ohio, 1966.

XVIII. Fluid Physics

SPACE SCIENCES DIVISION

A. The Stability of Viscous Three-Dimensional Disturbances in the Laminar Compressible Boundary Layer. Part II, L. M. Mack

The revision of the computer program to permit the solution of the eighth-order system of stability equations, which is described in SPS 37-45, Vol. IV, pp. 247-250, has been carried out. The revised program makes it possible to calculate the eigenvalues and eigenfunctions of three-dimensional as well as two-dimensional disturbances. It is the purpose of the present contribution to give a few results obtained from the eighth-order system, and, in particular, to establish the error involved in using the sixth-order system for three-dimensional disturbances.

The dissipation terms in the energy equation, when written in the tilde coordinate system, are

$$\begin{aligned} \tilde{D} = \gamma(\gamma-1) \frac{\tilde{M}_1^2}{\tilde{R}_\delta} & \left(2\mu \frac{du}{dy} \frac{\partial \tilde{u}'}{\partial y} + 2\mu \frac{du}{dy} \frac{\partial \tilde{v}'}{\partial \tilde{x}} \right. \\ & \left. + \frac{d\mu}{dT} \left(\frac{du}{dy} \right)^2 \frac{T'}{\cos^2 \psi} - 2\mu \frac{du}{dy} \frac{\partial \tilde{w}'}{\partial y} \tan \psi \right) \end{aligned}$$

In this coordinate system, the \tilde{x} axis is aligned with the wave normal, which is at angle ψ to the free-stream direction. The fluctuation velocity components \tilde{u}' and \tilde{w}' are

the components in the plane of the flow (x - z plane) parallel and perpendicular to the wave normal, and v' is the velocity component normal to the x - z plane. In the sixth-order system, the \tilde{z} momentum equation is omitted and the final term \tilde{D} , the only term in all of the equations where \tilde{w}' enters, must be neglected. Estimates presented in SPS 37-45 suggested that the sixth-order system could be used up to $M_1 = 3$ with an error of less than 10% in the amplification rate provided $R > 1000$ ($R = R_x^{1/2}$). This conclusion was based mainly on calculations made with $\tilde{D} = 0$ and the idea that the \tilde{w}' term should have about the same importance as the other terms combined. Unfortunately, these calculations contained a numerical error which became important for $M_1 > 3$. The correct result is that the four terms of \tilde{D} taken together have only a small effect on peak amplification rates at all Mach numbers for $R > 1000$. The effect of the fourth term alone does follow this general behavior, as is seen from Table 1. In this table, the amplification rates at a high and a low Reynolds number, at several Mach numbers, are given as computed from the sixth- and eighth-order systems. At each Reynolds number, the wave angle and wave number are near those of the most unstable disturbance. The choice of α at $M_1 = 10$ is arbitrary, since for a given ψ there is no amplification peak with respect to α in the first mode region. However, for $\alpha = 0.04$, $R = 1500$, the maximum amplification rate with respect to wave angle occurs at $\psi = 55$ deg.

Table 1. Comparison of amplification rates for three-dimensional disturbances as computed from sixth-order and eighth-order systems of equations

M_1	R	α	ψ , deg	Sixth order $\alpha c_i \times 10^3$	Eighth order $\alpha c_i \times 10^3$	Difference, %
1.3	500	0.075	45	0.883	0.824	7.2
1.3	1500	0.060	45	1.467	1.445	1.5
1.6	500	0.070	55	0.974	0.874	11.4
1.6	1500	0.050	55	1.384	1.346	2.8
2.2	500	0.055	60	1.198	1.066	12.4
2.2	800	0.045	60	1.391	1.300	7.0
2.2	1500	0.035	60	1.325	1.273	4.1
4.5	500	0.045	60	1.117	1.039	7.5
4.5	1500	0.050	60	1.641	1.613	1.7
5.8	500	0.050	55	0.790	0.736	7.3
5.8	1500	0.060	55	1.403	1.384	1.4
10.0	1500	0.040	55	0.444	0.434	2.3

No systematic checks of the differences between sixth- and eighth-order results have been made for values of α and ψ far from those of maximum amplification rates, or for other than first-mode disturbances. However, Table 1 certainly indicates that the sixth-order system is adequate for a great many three-dimensional calculations. This result is important because it is considerably more expensive to solve the eighth-order system than the sixth-order system. On the IBM 7094 Computer, it requires 0.14 s to integrate the three independent solutions of the sixth-order system across a single integration step. In contrast, it requires 0.25 s to similarly integrate the four independent solutions of the eighth-order system.

It is necessary to use the eighth-order system in order to obtain all of the eigenfunctions for a three-dimensional disturbance even when the eigenvalues can be computed from the sixth-order system. The eigenfunctions, or amplitude functions of p' , T' , v' and \tilde{u}' can be obtained from the eighth-order system. But to obtain the amplitude function, \tilde{h} , of \tilde{w} and hence the amplitude functions of the velocity components in any direction in the x - z plane, the eighth-order system is needed. As an example of the eigenfunctions of an unstable three-dimensional disturbance, the magnitudes of the amplitude functions \tilde{f} and \tilde{h} are shown in Fig. 1 and their phases in Fig. 2 for the $M_1 = 5.8$, $R = 1500$ disturbance of Table 1.

It is not obvious from the results in Figs. 1 and 2 that the $\partial \tilde{w} / \partial y$ dissipation term has such a small effect on the amplification rate as shown in Table 1. To obtain a

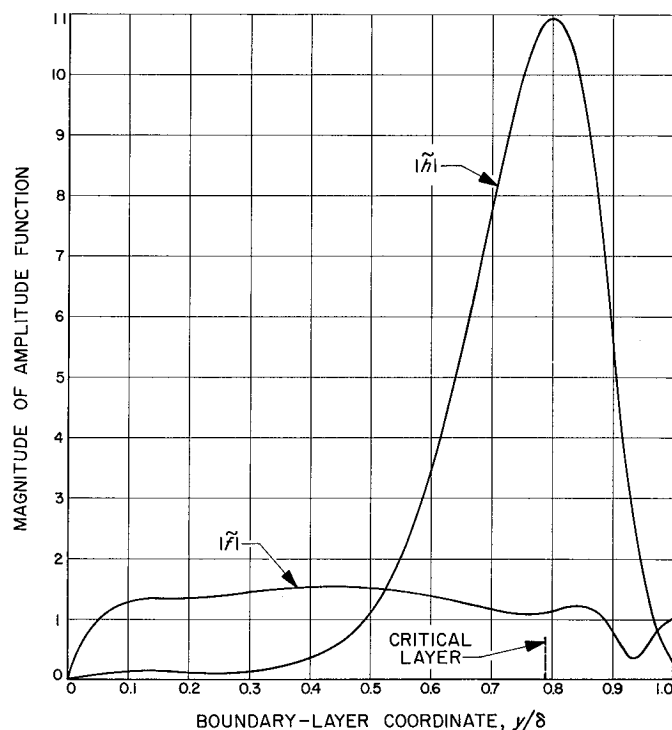


Fig. 1. Magnitudes of amplification functions \tilde{f} and \tilde{h} for 55-deg wave at $M_1 = 5.8$, $R = 1500$, and $\alpha = 0.060$ (insulated wall)

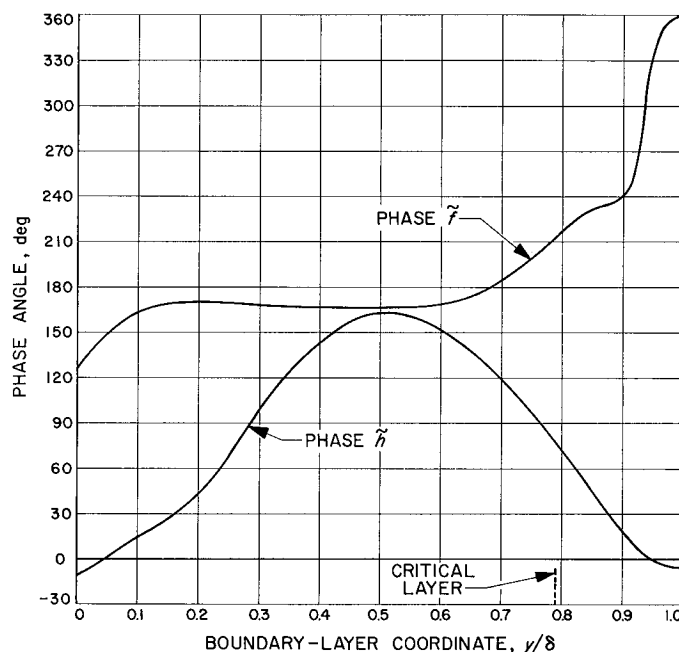


Fig. 2. Phase angles of amplification functions \tilde{f} and \tilde{h} for 55-deg wave at $M_1 = 5.8$, $R = 1500$, and $\alpha = 0.060$ (insulated wall)

Table 2. Effect of individual energy-equation dissipation terms on amplification rate at $R = 1500$ for three Mach numbers

Zero term	% change in αc_i		
	$M_1 = 2.2$	5.8	10.0
$2\mu \frac{du}{dy} \frac{\partial \tilde{u}'}{\partial y}$	+1.4	+0.7	+0.9
$2\mu \frac{du}{dy} \frac{\partial \tilde{v}'}{\partial x}$	0	-0.1	-0.4
$\frac{d\mu}{dT} \left(\frac{du}{dy} \right)^2 \frac{T'}{\cos^2 \psi}$	-0.1	-1.2	-2.4
$2\mu \frac{du}{dy} \frac{\partial \tilde{w}'}{\partial y} \tan \psi$	+4.1	+1.4	+2.3
All zero	+5.3	+0.8	+0.4

somewhat clearer idea of the effect of the individual dissipation terms on the amplification rate, calculations were made with each term set separately equal to zero. The results are shown in Table 2 for three Mach numbers. The disturbances are those given in Table 1 for $R = 1500$. The amplification rate depends linearly on each term, as shown by the fact that the change with $\tilde{D} = 0$, which was obtained by separate calculation, is just the sum of the individual changes. The $\partial \tilde{w}' / \partial y$ term is indeed the largest at all Mach numbers. However, its effect is less than three times the effect of the $\partial \tilde{u}' / \partial y$ term even though a \tilde{w}' gradient is present over more of the boundary layer than the \tilde{u}' gradient. The latter is concentrated near the wall, whereas the \tilde{w}' gradients are out near the critical layer where μ and du/dy are smaller than at the wall. Perhaps the main reason why the \tilde{w}' term fails to have a larger effect on the αc_i is that the effects of the positive and negative gradients tend to cancel out. Of the other two terms, the viscosity-fluctuation term is as important as the \tilde{w}' term at the two higher Mach numbers and has a destabilizing effect. The $\partial \tilde{v}' / \partial x$ term is only important at $M_1 = 10$.

The remaining calculation that has been made with the eighth-order system is a check with the available experimental amplification rates. Figure 3 gives a comparison of the theory with the measurements of Laufer-Vrebalovich (Ref. 1) at $M_1 = 2.2$, and of Kendall (Ref. 2) at $M_1 = 4.5$. The agreement is satisfactory in both cases, with the important caution at $M_1 = 2.2$ that the nature of the Laufer-Vrebalovich disturbances is unknown. If the disturbances were not essentially oblique waves with wave angles between about 40 to 70 deg, then the comparison is, of course, meaningless.

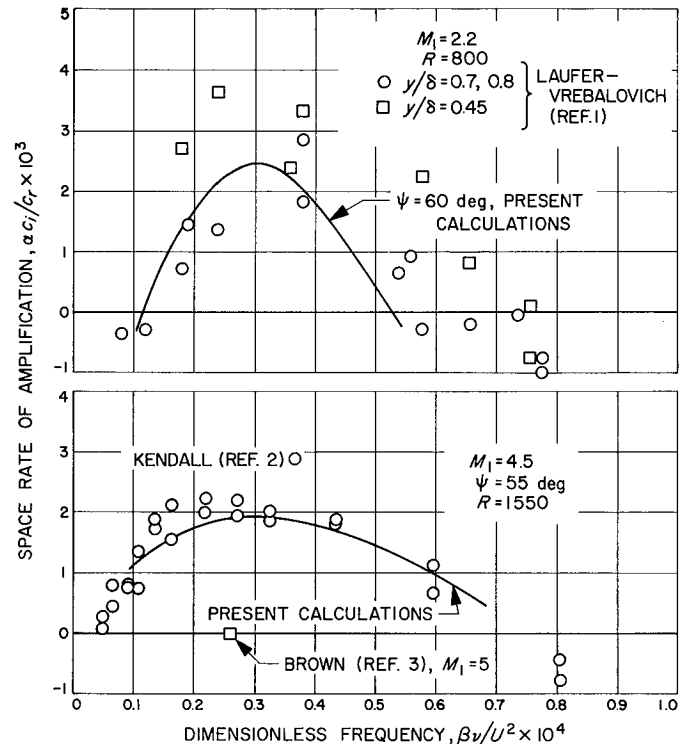


Fig. 3. Comparison of calculated amplification rates with measurements of Laufer-Vrebalovich and Kendall

An upper-branch neutral point computed by Brown (Ref. 3) at $M_1 = 5$ is also shown on Fig. 3. This point is for a 55-deg wave and was obtained from an eighth-order system of equations which includes all of the mean normal-velocity terms. These terms are neglected in the present calculations, which are based on the parallel-flow version of the stability theory. Brown found good agreement of his neutral-stability curve with the experimentally measured curve of Demetriades at $M_1 = 5.8$ (Ref. 4). It appears that if one accepts the Kendall experiment as decisive, then, in the absence of something dramatic occurring between $M_1 = 4.5$ and 5.8, the validity of the results of both Brown and Demetriades are suspect.

References

1. Laufer, J. and Vrebalovich, T., "Stability and Transition of a Supersonic Laminar Boundary Layer on an Insulated Flat Plate," *J. Fluid Mech.*, Vol. 9, Part 2, pp. 257-299, 1960.
2. Kendall, J. M., Jr., "Supersonic Boundary-Layer Stability," in *Proceedings of Boundary-Layer Transition Study Group Meeting*, Vol. II, U.S. Air Force Report BSD-TR-67-213, Vol. II, 1967.
3. Brown, W. B., "Stability of Compressible Boundary Layers," *AIAA J.*, Vol. 5, pp. 1753-1759, 1967.
4. Demetriades, A., "An Experiment on the Stability of Hypersonic Laminar Boundary Layers," *J. Fluid Mech.*, Vol. 7, Part 3, pp. 385-396, 1960.

XIX. Physics

SPACE SCIENCES DIVISION

A. Mechanism of the Reaction of Atomic Oxygen With Olefins, W. B. DeMore

1. Introduction

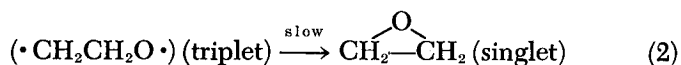
An earlier article (SPS 37-46, Vol. IV, pp. 212-217) described experiments on the addition of atomic oxygen to olefins dissolved in liquid argon at 87°K. The main purpose of the work was to measure activation energy barriers for O-atom addition to a series of olefins, but the results also provided new information on the mechanism of the reactions. The present article is a more detailed discussion from the latter point of view.

Addition of bivalent species such as O, S, and CH₂ to olefins has been a widely discussed topic in recent years. Much of the discussion has centered around the possible existence and properties of a biradical intermediate, which may be formed as follows:



By the rule of spin conservation, the unbonded electrons of $\cdot\text{CH}_2\text{CH}_2\text{O}\cdot$ must be unpaired; and, therefore, it has

often been assumed that rearrangement of the triplet biradical to give singlet products would be slow because of the requirement of spin inversion.



The suggestion has frequently been made that the triplet biradicals should be relatively long-lived intermediates in reactions of this type.

Although the importance of the Wigner-Witmer spin conservation rules is well established in simpler reactions involving atoms, there is little information available on the role of spin effects as rate-controlling factors in reactions involving more complicated molecules. Therefore, much of the interest in biradical chemistry has been based on the possibility of gaining more information on the actual effect of spin.

Much work has been done in efforts to identify and characterize biradicals, and even a brief review is impossible here. It may be said, however, that at the present

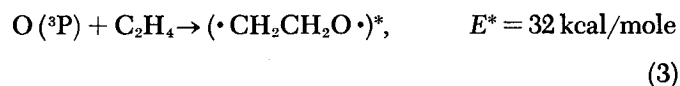
time the situation is somewhat unsatisfactory because a number of unanswered questions and contradictions remain. Biradicals have not been detected directly, and little is known about their chemical behavior or about the influence of spin state on the rates of their reactions. In addition, there seem to be inconsistencies in the reported behavior of biradicals formed by addition of O, S, and CH₂ to olefins, despite the fact that on general grounds one would expect analogous behavior for these species.

The present study does not prove or disprove the existence of biradicals, but does provide new limits on their behavior if it is postulated that they play a role in the addition of O-atoms to olefins.

2. Mechanism of the Reaction

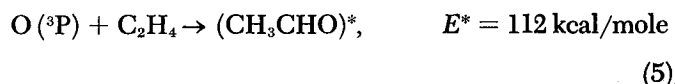
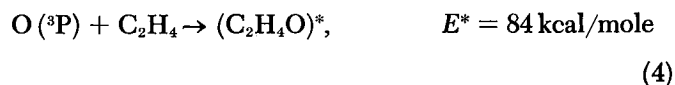
There are two properties of the experimental system which should be emphasized. First, because of the low temperature, the availability of thermal energy is quite low and the various reaction intermediates and products are energized almost entirely by reaction exothermicity. The second important property of the system is that deactivating collisions occur rapidly, and the rate of deactivation establishes a time reference against which other rates may be compared. On the basis of several observations, we find that moderately complex molecules are vibrationally deactivated in liquid argon at a rate of $10^{11.5 \pm 0.5} \text{ sec}^{-1}$. Therefore, unimolecular processes requiring the presence of vibrational energy do not occur in liquid argon unless the reaction rate is greater than $10^{11.5 \pm 0.5} \text{ sec}^{-1}$.

Taking an estimated value of 40 kcal/mole for the heat of formation of $\cdot\text{CH}_2\text{CH}_2\text{O}\cdot$ (Ref. 1), the initial step of biradical formation is exothermic by 32 kcal/mole.



Product analysis shows that isomerization of the biradical is not inhibited by the argon solvent, and that the product ratios are quite similar to those found in the gas phase. Now the isomerization reactions (ring closure and H-migration) undoubtedly have nonzero activation energies, and the activation energies almost certainly are not equal. Therefore, the biradical must undergo these isomerization reactions before any vibrational deactivation occurs, i.e., at a rate of about $10^{12.5 \pm 0.5} \text{ sec}^{-1}$.

The foregoing results show that the products ethylene oxide and acetaldehyde retain the full reaction exothermicity:



The behavior of the energized products can be calculated from the RRK equation.

$$k = A \left[1 - \frac{E_a}{E^*} \right]^{s-1} \text{ sec}^{-1} \quad (6)$$

The rate k is the unimolecular rate constant, and the quantities A and E_a are the pre-exponential factor and activation energy, respectively, from the Arrhenius equation $k = A \exp E_a/RT$. The quantity s is the number of effective oscillators of the molecule.

Results of the RRK calculations for $\text{C}_2\text{H}_4\text{O}^*$ and CH_3CHO^* are shown in Table 1. All the calculated rates are less than the collisional deactivation rate of $10^{11.5} \text{ sec}^{-1}$, so that the products are "frozen" in the initial distribution. This result confirms that the product CH_3CHO is not formed as a secondary product following $\text{C}_2\text{H}_4\text{O}$ isomerization, but instead is an initial product of the biradical rearrangement.

A composite mechanism is shown in Fig. 1, incorporating all the rate data. This mechanism accounts in a semi-quantitative way for the course of the $\text{O}(^3\text{P}) - \text{C}_2\text{H}_4$ reaction under conditions ranging from the gas phase at low pressures to the liquid phase experiments.

Table 1. Results of Rice-Ramsperger-Kassel calculations for reactions involved in $\text{O}(^3\text{P})$ addition to ethylene^a

Reaction	Rate parameters			log k , sec ⁻¹
	log A	E_a , kcal/mole	Source	
$\text{C}_2\text{H}_4\text{O}^* \rightarrow \text{CH}_3\text{CHO}^*$	14.5	57	Ref. 1	10.1
$\text{C}_2\text{H}_4\text{O}^* \rightarrow \cdot\text{CH}_2\text{CH}_2\text{O}\cdot^*$	—	—	— ^b	10.4
$\text{CH}_3\text{CHO}^* \rightarrow \cdot\text{CH}_2\text{CH}_2\text{O}\cdot^*$	13	85	— ^c	7.4
$\text{CH}_3\text{CHO}^* \rightarrow \text{CH}_3 + \text{CHO}$	15	79	Ref. 1	10.2

^a The value $s = 10$ was used for all calculations, based on thermal data as discussed in Ref. 1.

^b This rate was taken to be one-half the rate of isomerization to CH_3CHO .

^c The A -factor was estimated, and E_a was taken as equal to the endothermicity (80 kcal/mole) plus an approximate E_0 for the back reaction of 5 kcal/mole.

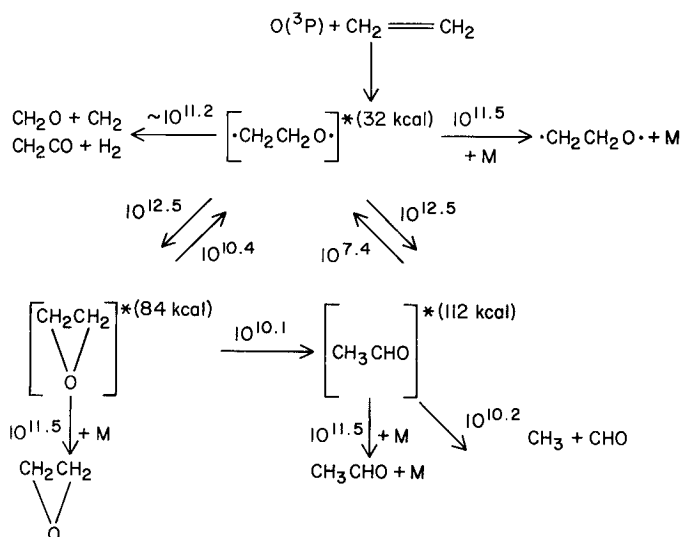


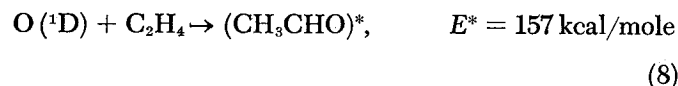
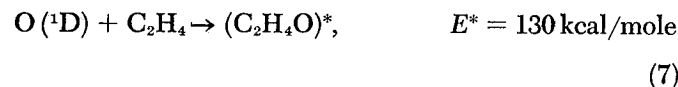
Fig. 1. Detailed mechanism for the reaction of O(³P) with ethylene

The most significant conclusions based on the mechanism of Fig. 1 are that at no point can any influence of the spin reversal requirement be detected, and that the correct results are obtained without taking any account at all of the spin question. Working backward from the RRK equation, it can be shown that the observed biradical isomerization rates of about $10^{12.5} \text{ sec}^{-1}$ correspond to A-factors of not less than about 10^{13} sec^{-1} , assuming that the activation energies are about 5 kcal/mole for those reactions. A value of 10^{13} sec^{-1} is a "normal" A-factor for reactions of this type, and certainly shows no indication of being spin-forbidden.

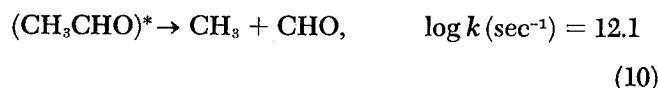
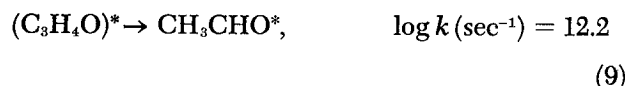
The most tenable explanation for the failure of spin reversal to be rate-controlling would be to postulate that the products are formed in triplet electronic states and subsequently are electronically deactivated. At the present time the only available information on the behavior of the products is in connection with the isomerization and decomposition rates. This data does not support the triplet state hypothesis, since the observed rates are just those that would be predicted for the ground states containing all the reaction exothermicity as vibrational energy. Although the triplet species might happen to decompose at the same rates as the singlet, it would seem rather coincidental that they should do so.

For the reaction of O(¹D) with C₂H₄, which was also studied, it is more difficult to formulate a detailed mechanism based on the product analysis. The reason is that

the electronic energy of O(¹D) appears as increased vibrational energy in the products.



The C₂H₄O* and CH₃CHO*, therefore, react more rapidly; the rates calculated by the RRK method are



Thus, an altered product distribution and a lower product yield are expected for the O(¹D) reaction. A change in product distribution cannot be measured experimentally because the initial distribution is not known. However, there is evidence for the occurrence of Reaction (10), since a decrease in product yield was observed for the O(¹D) reaction. This result tends to confirm the validity of the RRK calculations.

Reference

1. Benson, S. W., "Pyrolysis of Ethylene Oxide. A Hot Molecule Reaction," *J. Chem. Phys.*, Vol. 40, p. 105, 1964.

B. Dyadic Analysis of the World Models of

Cosmology, F. B. Estabrook, H. D. Wahlquist, and C. G. Behr¹

Cosmology has greatly advanced in recent years through (1) the discoveries of quasars and (2) the cosmic microwave blackbody radiation, and (3) by understanding the stellar processes and elemental abundances which result from nucleosynthesis. It appears that as this body of data and knowledge achieves some maturity as a science, a more sophisticated theory of the gravitation of the universe during its evolution may also be needed—and in particular, it may well be that simple Friedmann (expansion only) cosmological models may be inappropriate during the early stages, when the presence of rotation and shear could significantly change the time scales available for nuclear processes.

¹National Research Council postdoctoral resident research associate, supported by NASA.

Spatially homogeneous, but anisotropic, cosmological world models—solutions of the Einstein field equations of general relativity having not only expansion but also rotation and shear—have been considered principally by Taub (Ref. 1), and Heckmann and Schücking (Ref. 2). For the case of incoherent matter (dust), they derived a complicated system of six coupled second-order ordinary differential equations, together with some first-order equations which are integrals of these. Gödel (Ref. 3) announced without proof a Hamiltonian formulation of these, in a specialized case having local axial symmetry.

We have reformulated this entire problem in dyadic notation, with more general matter content—in particular, with perfect fluid; this enables the incorporation of pressure effects, which are important during the early stages. The utility of the dyadic formalism is well demonstrated; there results, in general, a set of nine first-order differential equations having (for the case of perfect fluid) two immediate first integrals. We are able to prove (what has previously been assumed) that the so-called Bianchi-Behr type of the cosmology is conserved, irrespective of the physical state of the matter content. We conveniently prove, and generalize, the theorem of Gödel. It appears that our first-order set is ideally suited for numerical calculation of the evolution of anisotropic world models, when combined with the first-order equations of nucleosynthesis.

A detailed account of this work is to be published (Ref. 4).

References

1. Taub, A., *Ann. Math.*, Vol. 53, p. 472, 1951.
2. Heckmann, O., and Schücking, E., *Gravitation: An Introduction to Current Research*, Chap. 11. Edited by L. Witten. John Wiley & Sons, Inc., New York, 1962.
3. Gödel, K., *Proc. Int. Cong. Math.*, Vol. I, pp. 175–181, 1950.
4. Estabrook, F. B., Wahlquist, H. D., and Behr, C. G., "Dyadic Analysis of Spatially Homogeneous World Models," *J. Math. Phys.*, Vol. 8, March 1968 (in press).

C. Unitary Representations of the Restricted Poincaré Group From a Unified Standpoint, J. S. Zmuidzinas and K. L. Phillips²

Our present-day understanding of the phenomena of elementary particle physics is based, in large measure, on the principles of special relativity. In fact, the quantum

mechanical states of a free particle transform according to a certain IUR of the restricted Poincaré group P_0 , the group of special relativity. The classification and construction of such representations is, therefore, of great importance in particle physics. The possible IURs of P_0 have been found by Wigner almost thirty years ago (Ref. 1). They fall into three general classes: those with m^2 (m = mass) positive, zero, and negative. Only the $m^2 \geq 0$ representations are believed to be physical. The $m^2 < 0$ representations would correspond to particles [tachyons (Ref. 2)] having a speed greater than the speed of light; so far they have not been observed. Moreover, the $m^2 < 0$ "particles" are mediators of forces between physical particles as, e.g., in the case of the exchange of a virtual (with $m^2 < 0$) pion between two colliding nucleons. It is, therefore, of great interest to investigate the IURs of P_0 falling into the third class, those with $m^2 < 0$.

Although all possible IURs of P_0 are known, as well as general methods of constructing them, there still remains the problem of constructing these IURs so as to have various desirable properties from the viewpoint of practical applications. In particular, it is desirable to construct IURs in terms of analytic functions (in the sense of complex variables) of the various parameters labeling the IURs, e.g., spin, helicity, and 4-momentum. This is a rather difficult technical problem, and hence only a brief outline of its proposed solution will be presented here. To make things more transparent, we shall discuss the familiar case of $SU(2, C)$, the covering group of the three-dimensional rotation group O_3^+ , drawing analogy, at the end, with the group of interest, P_0 . No proofs will be offered, the results for $SU(2, C)$ being classic.

As is well known, $SU(2, C)$ is a three-parameter Lie group of unitary 2×2 complex matrices under the usual matrix multiplication. An element $R(\alpha, \beta, \gamma)$ of $SU(2, C)$ may be written in the form

$$R(\alpha, \beta, \gamma) = R_3(\alpha) R_2(\beta) R_3(\gamma)$$

$$R_3(\alpha) = \begin{pmatrix} e^{-i\alpha/2} & 0 \\ 0 & e^{i\alpha/2} \end{pmatrix}$$

$$R_2(\beta) = \begin{pmatrix} \cos \beta/2 & -\sin \beta/2 \\ \sin \beta/2 & \cos \beta/2 \end{pmatrix}$$

where

$$0 \leq \alpha, \gamma < 2\pi$$

$$0 \leq \beta < \pi$$
(1)

²Consultant, Department of Mathematics, California Institute of Technology, Pasadena, Calif.

Let S be the unit sphere with points labeled by their polar angles θ and ϕ , where $0 \leq \theta \leq \pi$ and $0 \leq \phi < 2\pi$. The points on S are in a one-one correspondence with the unit vectors $\mathbf{n} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$ from the origin to the point (θ, ϕ) . By $L_2(S)$ we denote the Hilbert space of all μ -measurable complex-valued functions on S for which

$$\|f\| \equiv [\int |f(\mathbf{n})|^2 d\mu(\mathbf{n})]^{1/2} < \infty$$

where

$$d\mu(\mathbf{n}) \equiv \sin \theta d\theta d\phi$$

The inner product on $L_2(S)$ is just

$$(f, g) = \int \overline{f(\mathbf{n})} g(\mathbf{n}) d\mathbf{n} \quad [f, g \in L_2(S)]$$

Given an $R \in SU(2, C)$, put

$$R: f(\mathbf{n}) \rightarrow (T_R f)(\mathbf{n}) \equiv f(R^{-1}\mathbf{n})$$

where $R^{-1}\mathbf{n}$ is the unit vector obtained from \mathbf{n} by the rotation R^{-1} . One verifies that $f(\mathbf{n}) \rightarrow f(R^{-1}\mathbf{n})$ is a representation of $SU(2, C)$ and, because of $d(R\mathbf{n}) = d\mathbf{n}$, that it is even unitary. In general, it will not be irreducible, unless the function f is suitably restricted. Functions transforming irreducibly are well known. They are simply the rotation matrices

$$D_{\mu s}^s(\mathbf{n}) = D_{\mu s}^s(\phi, \theta, 0) \equiv \phi_{s\mu}(\mathbf{n})$$

Here s and μ are elements of what one may call "the space of IUR labels," denoted by S^* :

$$S^* = \{(s, \mu); s = 0, 1/2, 1, 3/2, \dots; \\ \mu = s, s-1, \dots, -s+1, -s\}$$

The space S^* is in a sense dual to S , as the notation implies. Namely, S^* is the space of basis functionals on the space S , the number of parameters in both S and S^* being the same, namely two. The functions $\phi_{s\mu}(\mathbf{n})$ are just the transformation coefficients between the two spaces. An arbitrary function in $L_2(S)$ can be expressed as a linear combination of the $\phi_{s\mu}(\mathbf{n})$. We finally remark that S , as a manifold, is a submanifold of the group manifold $SU(2, C)$. The latter consists of all triples (α, β, γ) subject to Eq. (1), while the former obeys the additional restriction $\gamma = 0$. In summary, for $SU(2, C)$ we have the following main facts:

- (1) The "representation manifold" S is a submanifold of the group manifold.

- (2) Unitary representations of $SU(2, C)$ are realized in terms of functions in $L_2(S)$.
- (3) An arbitrary unitary representation of $SU(2, C)$ in $L_2(S)$ may be decomposed into IURs, i.e., written as a linear combination of the $\phi_{s\mu}(\mathbf{n})$.
- (4) The space of IUR labels, S^* , is dual to S and has the same number of parameters as S .

The same sort of situation holds for the restricted Poincaré group P_0 . Thus, the representation manifold M (six-dimensional) may be chosen as a submanifold of P_0 (ten-dimensional). Because of the noncompact nature of P_0 , it is impossible to realize unitary representations of P_0 in terms of square-integrable functions: one must use generalized functions or distributions. Similarly, decomposition of an arbitrary unitary representation of P_0 into IURs is to be understood in the sense of distributions. Construction of functions analogous to the $\phi_{s\mu}(\mathbf{n})$ is not trivial and will not be described. The dual M^* of M has a rather complicated structure due to the fact that the three classes of IURs of P_0 have basically different mathematical properties. The three classes of IURs are constructed on functions defined on three separate submanifolds of M , so that M can be written as $M = M_+ \cup M_0 \cup M_-$ in an obvious notation. Corresponding to this, one has $M^* = M_+^* \cup M_0^* \cup M_-^*$ and the dual relations $M_\pm \leftrightarrow M_\pm^*$ and $M_0 \leftrightarrow M_0^*$. It will be shown elsewhere that for M one may take the topological product of E_4 (euclidean manifold in four-dimensions) and the complex z -plane. Then $M_+ = E_4 \times \{z: |z| < 1\}$, $M_0 = E_4 \times \{z: |z| = 1\}$, and $M_- = E_4 \times \{z: |z| > 1\}$. Analytic continuation of the IUR functions in the parameters of M^* presents no particular difficulty as long as one stays within each of the three submanifolds of M^* . Continuation from one submanifold to another is, however, a delicate matter and is still under investigation.

References

1. Wigner, E. P., *Ann. Math.*, Vol. 40, p. 149, 1939.
2. Feinberg, G., *Phys. Rev.*, Vol. 159, p. 1089, 1967.

D. Testing Analytic Models Against Compressed Spectral Data, E. L. Haines, R. H. Parker,³ and R. Gouw⁴

In any experiment where several parameters of a system are measured repeatedly, a large volume of data may be

³Department of Physics, College of William and Mary, Williamsburg, Virginia.

⁴Informatics, Inc., Los Angeles, California.

accumulated. This data may be made more manageable by the use of data compression techniques which preserve the significant information and discard that which is statistically insignificant. The final step in the analysis is to interpret the compressed data in terms of a mathematical model or hypothesis for the system. This summary describes a method for fitting the data to models and determining whether the model adequately describes this data.

This method has been developed to aid in analyzing secondary electron yield data recorded as a function of two parameters which describe the primary ion, a fission fragment passing through a metal film (Ref. 1). Analysis in the computer consisted of transforming the observed quantities, sorting the events according to mass and energy of the fragment, and compressing the spectrum of electron yield (Refs. 2, 3, and SPS 37-42, Vol. IV, p. 167). The result of this analysis was a compressed electron spectrum for each combination of mass and energy. Compressed data took the form of five coefficients of the Gauss-Hermite expansion. How well these coefficients described the spectra of electron yields may be seen in Ref. 3.

The choice of compression mode, i.e., the Gauss-Hermite coefficients, was natural for these single peaked spectra, because each coefficient represented some physical aspect of the monomodal spectrum. For example, the zero order coefficient represented the area of the spectrum, the first order represented the mean, second order the variance, third order the skewness, etc. No further reduction of the coefficients was necessary, because these coefficients, particularly the mean and variance, could be analyzed in terms of an analytic model.

Frequently, this finite sequence of terms in the expansion is not adequate. For instance, a model consisting of a Gaussian peak on a linear background may be more appropriate to describe the physical system. Another simple example is two overlapping Gaussians. In these cases the step from the compressed data to the model is more complicated. The discussion which follows presents a general approach to solving this problem.

In order to attain any degree of compression, the number of coefficients used to describe the spectrum must be smaller than the number of subdivisions or channels used to collect the spectrum. The coefficients are derived from the spectrum, $f(x)$, by the transform

$$c_i = Q_i^{-1} \int_{D_x} f(x) h(x) P_i(x) dx, \quad i = 0, 1, \dots, r \quad (1)$$

where

$$Q_i = \int_{D_x} h(x) P_i^2(x) dx \quad (2)$$

In Eqs. (1) and (2), D_x is the domain of x for which the set of polynomials P_i is orthogonal. The polynomials are orthogonal with respect to the so-called weight function $h(x)$. Let us introduce an analytic model of the physical process $g(x; \alpha_1, \alpha_2, \dots, \alpha_p)$ (there may be a variety of applicable models) whose unknown parameters are $\alpha_1, \alpha_2, \dots, \alpha_p$. If $r > p$ (the number of coefficients is greater than the number of parameters), it should be possible to solve some set of equations for these parameters.

First, the model must be in a form which can be compared with the coefficients. This is accomplished by transforming the model in a manner analogous to Eq. (1),

$$\phi_i = Q_i^{-1} \int_{D_x} g(x; \alpha_1, \alpha_2, \dots, \alpha_p) h(x) P_i(x) dx, \quad i = 0, 1, \dots, r \quad (3)$$

The quantities ϕ_i are model coefficients. The comparison residue is given simply by

$$\delta_i = \phi_i - c_i \quad (4)$$

The problem is to solve for the parameters of the model in some manner that will lead to their best statistical estimates, a manner that will reduce the set of residues δ_i to a minimum in a minimum-variance sense. The method of least squares cannot be applied because it assumes that the observables, in this case the coefficients, are uncorrelated. The method applicable here is a more general method called minimum-variance linear unbiased estimation (Refs. 4, 5, and 6). It too is based on a test of chi-square, where chi-square is calculated by the double summation

$$\chi^2 = \sum_i \sum_j \lambda_{ij} \delta_i \delta_j \quad (5)$$

In matrix notation, this becomes

$$\chi^2 = \Delta^\dagger \Lambda \Delta$$

where Λ is the "weight" matrix, a symmetrical matrix of rank r , and Δ is a vector of length r .⁵ The superscript \dagger denotes the transpose of a vector or matrix. Chi-

⁵The "weight" or information matrix as normally used is the inverse of the covariance matrix of the observations. Here information theory is used to calculate the matrix directly, using methods derived from R. A. Fisher (Ref. 7).

square is minimized with respect to variation of each of the parameters. This is accomplished by setting the differentials of chi-square with respect to each parameter equal to zero, giving the matrix equation

$$\frac{\partial \chi^2}{\partial \alpha_p} = -2 \left(\frac{\partial \Delta}{\partial \alpha} \right)^{\dagger} \Lambda \Delta = 0 \quad (6)$$

In the general case, $g(x; \alpha_1, \alpha_2, \dots, \alpha_p)$ may have non-linear dependence on some or all of its parameters, α , yielding Eq. (6) insoluble. However, the equation may be solved numerically by applying a sequence of linear adjustments to the parameters. The vector Δ is expanded in a truncated Taylor's series in the parameters thus,

$$\Delta = \Delta_0 + \frac{\partial \Delta}{\partial \alpha} \delta \alpha \quad (7)$$

where $\delta \alpha$ represents a small linear change in the set of parameters. Let us define the rectangular $r \times p$ differential matrix by the symbol A ,

$$\frac{\partial \Delta}{\partial \alpha} = A \quad (8)$$

and substitute Eq. (7) and Eq. (6). The result is a set of normal equations

$$A^{\dagger} \Lambda \Delta_0 = -(A^{\dagger} \Lambda A) \delta \alpha \quad (9)$$

This has the simple form

$$z = \Omega x \quad (10)$$

This is easily solved for $\delta \alpha$ [x in Eq. (10)].

A computer program was written which performed the following operations:

- (1) Calculated Λ , the "weight" matrix, using equations derived from information theory (Ref. 7).
- (2) Using estimates for the model's parameters, evaluated the matrix A and vector Δ_0 .
- (3) Calculated χ^2 .
- (4) Solved Eq. (9) for $\delta \alpha$.
- (5) Altered the model's estimated parameters, α , by the amounts $\delta \alpha$.
- (6) Repeated (2) through (5) until χ^2 reached a minimum.

- (7) Calculated statistical quantities related to χ^2 which help judge the validity of the model. Printed the model's parameters and their standard deviations, and provided graphical output.

Results from this computer program are shown in the figures that follow. Mock experiments were performed by the Monte Carlo method, experiments which use well-defined analytic functions but which introduce statistical fluctuations in the same manner as real experiments. The functions used in the Monte Carlo generator were overlapping Gaussian peaks; these or similar overlapping peaks are often encountered in physical or chemical spectra. Compression was performed in terms of Chebyshev polynomials or Fourier series. The results shown here are chosen from spectra compressed as Fourier coefficients.

Figure 2 compares tests of two analytic models (hypotheses) against the same data. The Monte Carlo experiment defined three peaks, the center of which represented only 5% of the spectrum's area. It was located in the tail of the larger peak. The spectrum was compressed *during its collection* (Ref. 3 and SPS 37-42, Vol. IV, p. 167) into 16 Fourier coefficients. The expansion of the 16 terms is represented by the plus marks. The "ringing" of the right-hand end of the spectrum reveals that the 16 coefficients were not adequate to describe the spectrum completely. However, it is emphasized that this did not impair the validity of the tests, because the models were treated in the same manner. That is, they were transformed into 16 model coefficients. The two models tested were: (1) "There are only two peaks here" (Fig. 2a) and (2) "There are three peaks here, one at about channel 36 which is not as obvious as the other two" (Fig. 2b). Written in analytic form and treated by the program, these models yielded the continuous lines in the two graphs of Fig. 2. Chi-square tests showed model 1 to be invalid, and model 2 to be acceptable. For model 2, the program gave back model parameters which located the small peak and gave its area and variance. These parameters compared within their statistical deviations with those parameters used in the Monte Carlo generator.

Figure 3 exhibits the results of a much more exacting pair of tests. The Monte Carlo generator again created a spectrum containing three peaks, but each was 50% wider than in the previous test, and the small central peak contained only 2% of the spectral area. The expansion of the 16 Fourier coefficients, the "plus" marks, revealed nothing of the third peak. Yet when the program tested the same two models, two peaks or three, the two-peak test was rejected absolutely (Fig. 3a), while the

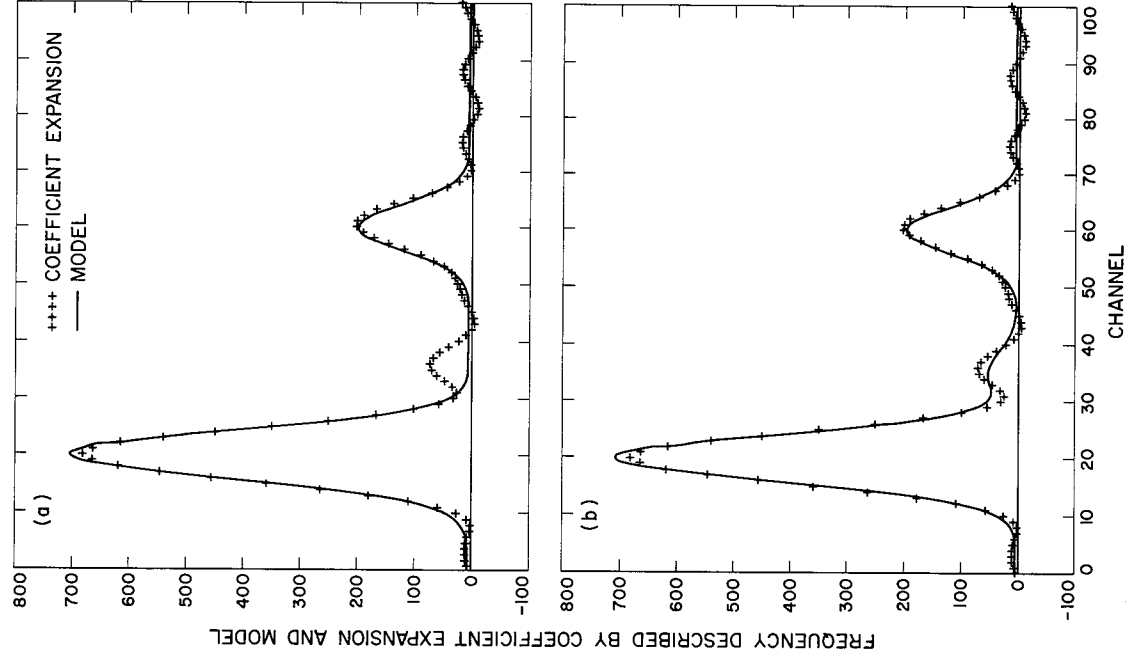


Fig. 2. Results of tests of (a) two-peak and
(b) three-peak analytic models against
one set of compressed data

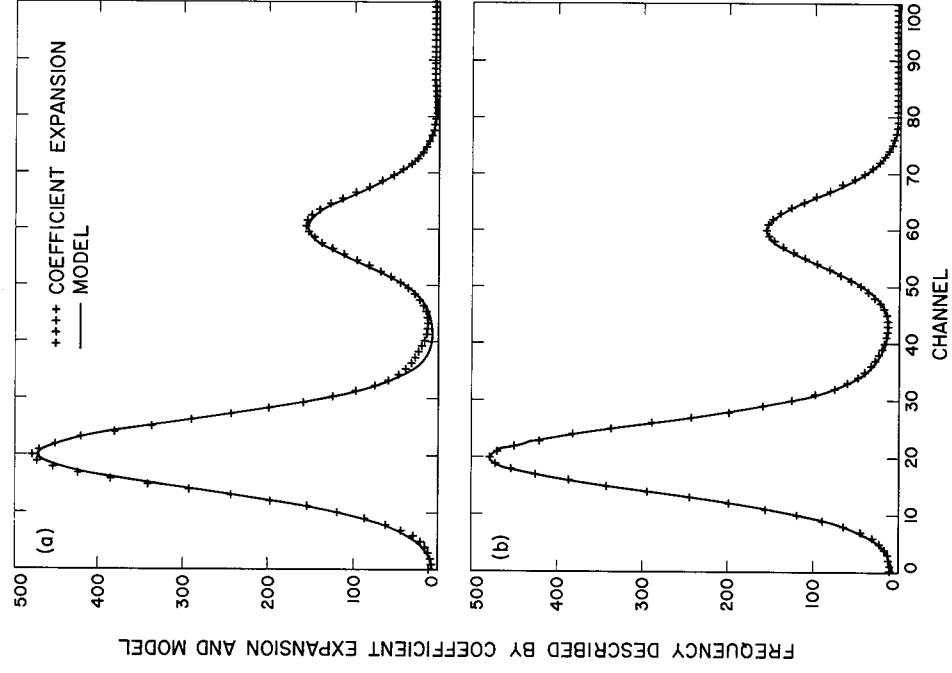


Fig. 3. Results of tests of (a) two-peak and
(b) three-peak models

three-peak test was found acceptable (Fig. 3b). However, the parameters of the hidden peak in the three-peak test were not accurately reproduced. Thus, while the program answered important qualitative questions about the data, it could not accurately locate the peak.

Figure 4 presents plots of data coefficients and the model coefficients, c_i and ϕ_i , respectively, against the order of the coefficients for the two tests shown in Fig. 3. It was these coefficients which were treated in the fitting process. Small differences between the coefficients, data and model, were the δ_i which together made up the vector Δ . When the vector Δ was large, some or all of the coefficients showed separation as in Fig. 4a. When Δ was small, the coefficients nearly coincided, as in Fig. 4b. Small Δ corresponded to small χ^2 , indicating good agreement between the data and the model.

A variety of tests have been performed on other multiple-peaked distributions using both Chebyshev compression and Fourier compression. In one case the results of a test on compressed data were directly compared with the results of a normal least-squares test on the uncompressed data. The two tests yielded similar results in the chi-square test of validity, and the same values for the model's parameters within standard deviation. Even the standard deviations given by the two methods were nearly the same. This clearly demonstrated that all of the "significant" information of the spectrum was contained in the 16 data coefficients, and that the method described in this article extracted the information as efficiently as the method of least squares.

It is worth noting that this method of testing analytic models is applicable to any experiment, provided orthogonal polynomials are used in the compression scheme (Ref. 6). The method is equally applicable to earth-bound data compressed to save computer memory space and to compressed data taken aboard a planetary lander or a distant spacecraft.

References

1. Haines, E. L., and Whitehead, A. B., "Secondary Electron Emission Induced by the Passage of Fission Fragments through Metal Foils," *Bull. Am. Phys. Soc.*, Vol. 12, No. 2, p. 208, 1967.
2. Haines, E. L., "The Use of Local Linear Transforms to Reduce the Size of a Two-Dimensional Associative Memory," *IEEE Trans. of the Nuclear Science Symposium* (to be published).
3. Whitehead, A. B., Parker, R. H., and Haines, E. L., "Use of Transformations in Multiparameter Data Sorting," *IEEE NS-14*, No. 1, p. 599, 1967.

4. Plackett, R. L., *Regression Analysis*, pp. 31-51. Clarendon Press, Oxford, 1960.
5. Scheffé, H., *The Analysis of Variance*. John Wiley & Sons, Inc., New York, 1959.
6. Kizner, W., *The Enhancement of Data by Data Compression Using Polynomial Fitting*, Technical Report 32-1078. Jet Propulsion Laboratory, Pasadena, Calif., Oct. 1, 1967.
7. Fisher, R. A., *Statistical Methods and Scientific Inference*, pp. 141-175. Hafner Publishing Co., New York, 1959.

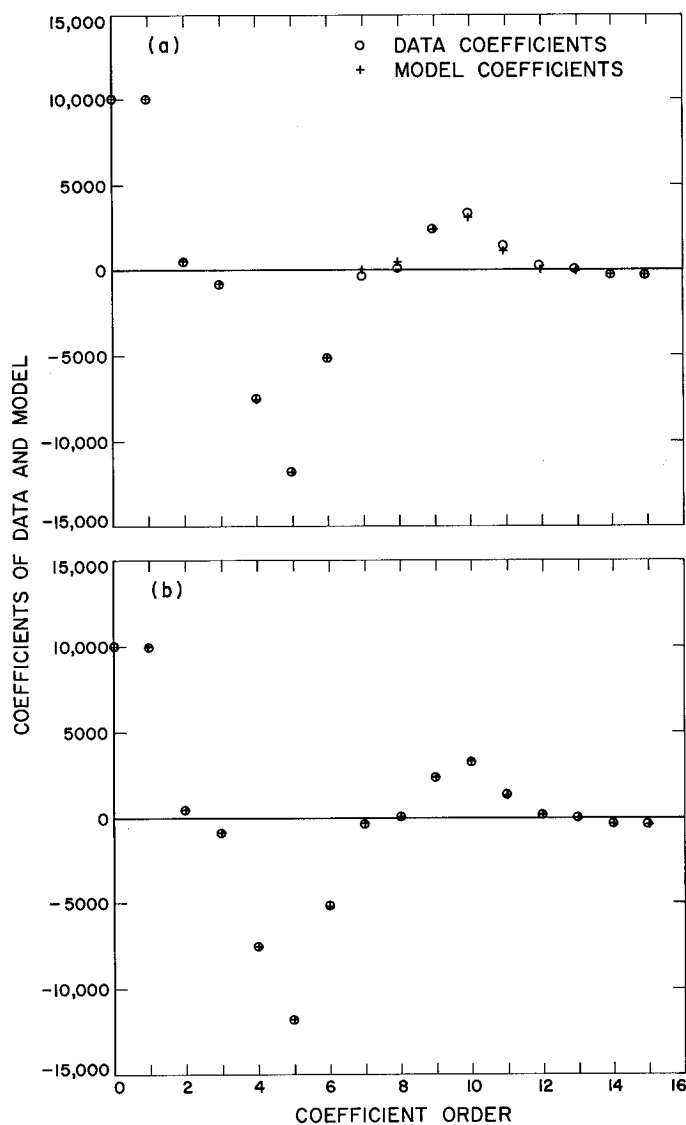


Fig. 4. Plot of data coefficients and model coefficients against coefficient order for (a) two-peak and (b) three-peak cases illustrated in Fig. 3

PRECEDING PAGE BLANK NOT FILMED

XX. Communications Systems Research

TELECOMMUNICATIONS DIVISION

A. Sequential Decoding With Decision-Directed Phase Estimation, J. A. Heller

1. Introduction

A method of communicating coded information over a channel with a time-varying phase is described and analyzed here. All of the available transmitter power is used in the data signal. Channel phase measurement based on the biphase-modulated data signal is accomplished by using past channel symbol decisions to effectively remove the modulation. Results in the form of bounds on the required energy per bit for reliable communication are obtained as a function of the signal energy-to-noise ratio in the bandwidth of the unmodulated received signal.

Suppose one of M messages m_i is chosen for transmission by the source. The transmitted signal of duration T will have the form

$$\begin{aligned} s_i(t) &= \left[\sum_{k=1}^N a_{ik} b\left(t - k \frac{T}{N}\right) \right] (2S)^{1/2} \cos \omega_0 t \\ &= \sum_{k=1}^N s_{ik}(t) (2)^{1/2} \cos \omega_0 t \end{aligned} \quad (1)$$

where $b(t)$ is a unit energy modulating waveform; it is non-zero only for $-T/N \leq t \leq 0$. The a_{ik} 's are either $+1$ or -1 , depending on the N element binary code word that specifies $s_i(t)$. S is the average transmitter power. Thus $s_i(t)$ is made up of a sequence of N non-overlapping binary waveforms or channel symbols, of duration T/N . The channel corrupts the signal by adding white Gaussian noise and imposing a time-varying phase shift $\theta(t)$ on the carrier. It is assumed that $\theta(t)$ does not vary significantly in the time duration of a channel symbol (T/N seconds). The received signal during the k th of the N signaling intervals is thus

$$r_k(t) = a_{ik} b\left(t - k \frac{T}{N}\right) (2S)^{1/2} \cos(\omega_0 t - \theta_k) + n_k(t) \quad (2)$$

Demodulation is accomplished by passing $r_k(t)$ through filters matched to the in-phase and quadrature components of the transmitted signal. The filter outputs are sampled every T/N seconds in synchronism with the end of each signaling interval (bit sync is assumed). This results in a pair of observables for each signaling interval

$$\begin{aligned} r_{ik} &= a_{ik} [(E_N)^{1/2} \cos \theta_k + n_{ck}] \\ r_{sk} &= a_{ik} [(E_N)^{1/2} \sin \theta_k + n_{sk}] \end{aligned} \quad (3)$$

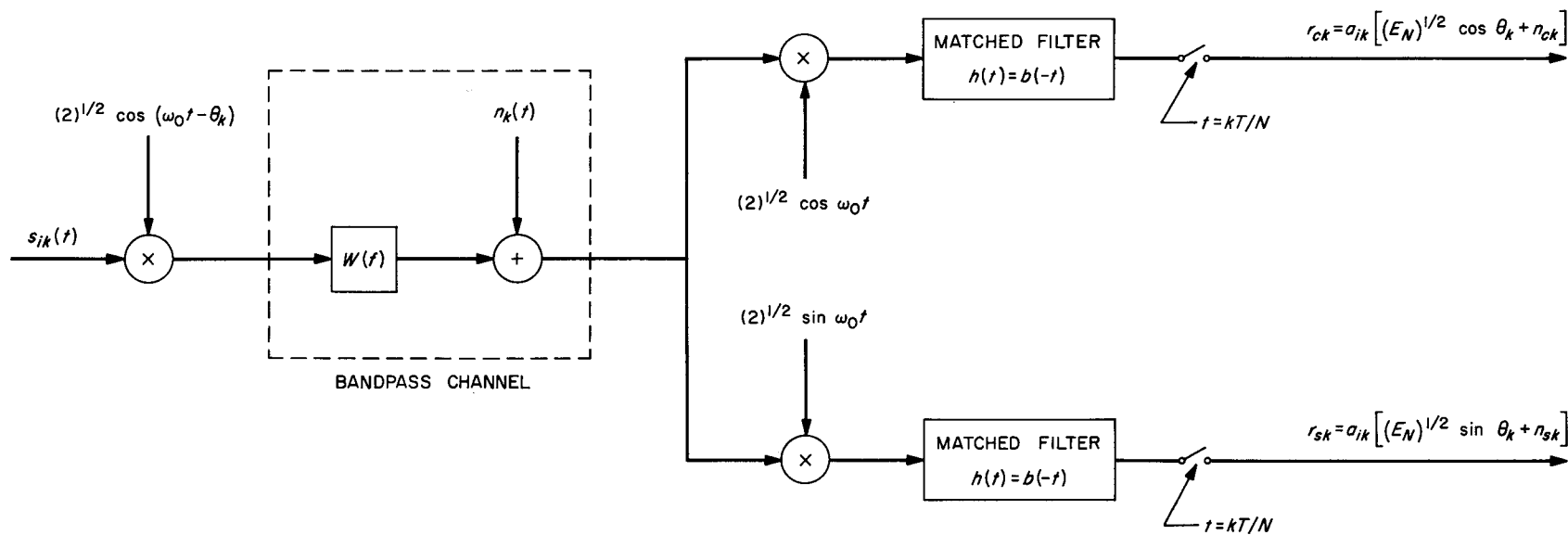


Fig. 1. Channel and demodulator

Here $E_N = ST/N$ is the energy per channel symbol and the n_{ck} 's and n_{sk} 's are statistically independent, zero mean Gaussian random variables with variance $N_0/2$. The effects of the channel and demodulator are summarized in Fig. 1. In this figure $W(f)$ is a bandpass filter centered at ω_0 ; it represents the bandpass channel. It can be shown (Ref. 1) that the demodulator destroys no information relevant to optimum detection.

Now, suppose an estimate of the phase in the k th interval $\hat{\theta}_k$ is desired. For this purpose it is convenient to consider the pair of observables in Eq. (3) as a normalized complex random variable

$$R_k = a_{ik} z_k = \frac{1}{(N_0)^{1/2}} (r_{ck} + j r_{sk}) \quad (4)$$

where

$$\bar{z}_k = \left(\frac{E_N}{N_0} \right)^{1/2} \exp(j\theta_k), \quad \sigma_{z_k}^2 = 1$$

In the absence of noise, the angle of z_k is θ_k ; therefore, if \hat{a}_{ik} is an estimate of a_{ik} , one estimate of θ_k would be

$$\hat{\theta}_k = \text{angle of } (\hat{a}_{ik} R_k) = \text{angle of } (\hat{a}_{ik} a_{ik} z_k) \quad (5)$$

If the modulation estimate is correct ($\hat{a}_{ik} = a_{ik}$), then $\hat{\theta}_k = \text{angle of } (z_k)$. This is shown geometrically in Fig. 2. Here R_k is multiplied by \hat{a}_{ik} in an attempt to remove the ± 1 modulation which otherwise would introduce a 180-deg phase ambiguity. If $\theta(t)$ changes slowly enough, the estimate could be improved by observing the received

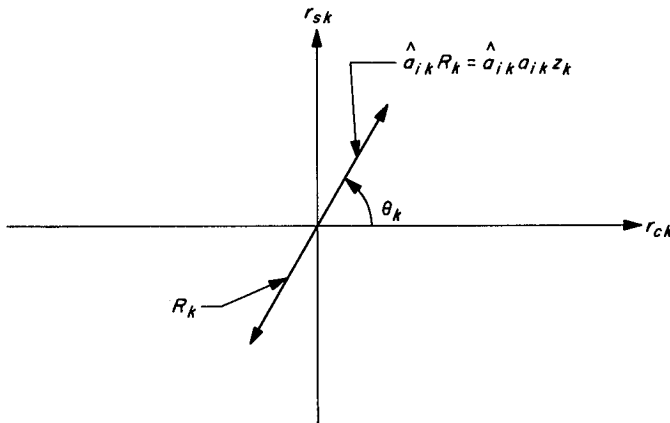


Fig. 2. Resolution of 180-deg phase ambiguity using modulation estimates (in this case $\hat{a}_{ik} = a_{ik} = -1$)

data in several preceding signaling intervals

$$\hat{\theta}_k' = \text{angle of } (Q_k) \quad (6a)$$

where

$$Q_k = \sum_{l=k-p}^{k-1} \hat{a}_{il} R_l \quad (6b)$$

Equation (6) is an estimate of the channel phase in interval k based on the received data, and the modulation decisions from the past p intervals. This method of phase estimation has been called "decision-directed channel phase measurement" (Ref. 2).

If the past modulation decisions \hat{a}_{il} , $k-p \leq l \leq k-1$, are correct, and if the phase changed a relatively small amount over the past p signaling intervals, then Q_k will tend to be at an angle near θ_k as shown in Fig. 3. Now if it is hypothesized that $\hat{a}_{ik} = a_{ik}$, then from Eq. (4) $\hat{a}_{ik} R_k$ will tend to be at angle θ_k if the hypothesis is correct, and at angle $\pi + \theta_k$ if it is incorrect. The dot-product of the two-dimensional vectors specified by Q_k and $\hat{a}_{ik} R_k$ in Fig. 3 is thus a measure of how likely it is that \hat{a}_{ik} is correct. Figure 3 shows a case where it is likely that a_{ik} is correct; that is, the dot-product is positive. In terms of the complex numbers involved, the dot-product w_{ik} is

$$\begin{aligned} w_{ik} &= \frac{1}{2} \hat{a}_{ik} (Q_k R_k^* + Q_k^* R_k) \\ &= \frac{1}{2} \sum_{l=k-p}^{k-1} \hat{a}_{ik} a_{il} \hat{a}_{il}^* a_{il} (z_l z_k^* + z_l^* z_k) \end{aligned} \quad (7)$$

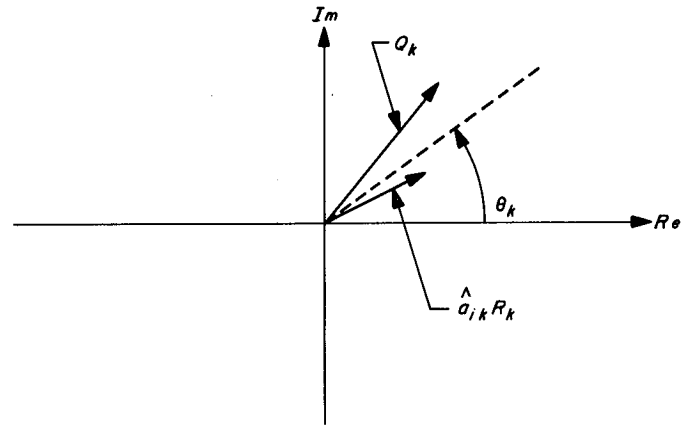


Fig. 3. Channel reference vector Q_k and received data vector R_k , premultiplied by modulation hypothesis \hat{a}_{ik}

In certain limiting cases it can be shown (Ref. 1) that w_{ik} is actually monotonic with the likelihood of \hat{a}_{ik} given the past received data and modulation.

2. Block Code Probability of Error

Suppose initially that one of two possible messages is to be sent. The signals corresponding to the code words are of the form shown in Eq. (1). For a given modulating waveform $b(t)$, the sets $\{a_{1k}\}$ and $\{a_{2k}\}$, $1 \leq k \leq N$, completely specify the signals corresponding to messages m_1 and m_2 , respectively. For the purposes of analysis it is assumed that the code letters a_{ik} are chosen at random with equal probability of being $+1$ or -1 , each a_{ik} being independent of all others for $1 \leq k \leq N$, $i = 1, 2$.

Suppose message m_1 is sent. Decoding proceeds as follows: the received data R_k is first tested against $s_1(t)$ by forming the dot-product w_{ik} for all k , $1 \leq k \leq N$, and summing

$$\begin{aligned} t_1 &= \sum_{k=1}^N w_{1k} \\ &= \frac{1}{2} \sum_{k=1}^N \sum_{i=k-p}^{k-1} \hat{a}_{1k} a_{1k} \hat{a}_{1i} a_{1i} (z_i z_k^* + z_i^* z_k) \end{aligned} \quad (8)$$

Note that a_{1i} is used in the above expression because, in fact, m_1 was sent. The \hat{a}_{1i} are simply chosen as the code letters of the message under test; so, for message 1

$$\hat{a}_{1i} = a_{1i} \quad (9)$$

and

$$\begin{aligned} t_1 &= \frac{1}{2} \sum_{k=1}^N \sum_{i=k-p}^{k-1} a_{1k}^2 a_{1i}^2 (z_i z_k^* + z_i^* z_k) \\ &= \frac{1}{2} \sum_{k=1}^N \sum_{i=k-p}^{k-1} (z_i z_k^* + z_i^* z_k) \end{aligned} \quad (10)$$

This will tend to be a large positive number. For, since m_1 was sent and the received data is presently being compared to $s_1(t)$, all of the decisions \hat{a}_{1i} are correct. Therefore, each dot-product w_{1k} in the summation comprising t_1 tends to be positive. Note that for channel symbols near the beginning of the code word ($k < p$), a p channel symbol reference is not possible because there are less than p previous symbols available (i becomes

negative in the inner summation in Eq. 8). This can be remedied by transmitting a p symbol known "initializing reference" prior to code word transmission.

The same procedure is used in comparing the received data to $s_2(t)$. This results in

$$t_2 = \frac{1}{2} \sum_{k=1}^N \sum_{i=k-p}^{k-1} a_{2k} a_{1k} a_{2i} a_{1i} (z_i z_k^* + z_i^* z_k) \quad (11)$$

Because of the code ensemble used to generate the a_{ik} 's, the coefficient $a_{2k} a_{1k} a_{2i} a_{1i}$ is $+1$ or -1 with equal probability; hence, the average value of t_2 is zero. The receiver now chooses that m_i for which t_i is largest, $i = 1, 2$; an error occurs if $t_2 \geq t_1$. Again taking into account the code statistics, the two-signal error probability is

$$\begin{aligned} P_2(\epsilon) &= \Pr [t_1 - t_2 \leq 0] \\ &= \Pr \left[\sum_{k=1}^N \sum_{i=k-p}^{k-1} b_{ki} (z_i z_k^* + z_i^* z_k) \leq 0 \right] \end{aligned} \quad (12)$$

where the b_{ki} 's are 0 or 1, depending on whether $a_{2k} a_{1k} a_{2i} a_{1i}$ is $+1$ or -1 . For fixed b_{ki} 's (fixed code), the two-signal error probability is therefore the probability that a quadratic form in complex Gaussian random variables is less than zero. Normalization by dividing b_{ki} by p results in

$$P_2(\epsilon) = \Pr [\mathbf{Z}^{T*} \mathbf{B} \mathbf{Z} \leq 0] \quad (13)$$

Here \mathbf{Z} is a column vector of the elements z_k , and \mathbf{B} is an $N \times N$ symmetric matrix. All elements of \mathbf{B} must be zero except those on the p diagonals directly above and below the principal diagonal. These elements are either 0 or $1/p$, depending on whether b_{ki} is 0 or 1.

By using a Chernoff bound, the probability in Eq. (13) can be bounded as (Ref. 3)

$$P_2(\epsilon) \leq \min_{0 \leq \eta \leq 1} \prod_{m=1}^N \frac{1}{1 + \eta \lambda_m} \exp \left(- \frac{\eta \lambda_m}{1 + \eta \lambda_m} |d_m|^2 \right) \quad (14)$$

where the λ_m 's are the eigenvalues of \mathbf{B} and the d_m 's are the projections of the mean of \mathbf{Z} onto the eigenvectors of \mathbf{B} .

$$d_m = \mathbf{e}_m^* \cdot \bar{\mathbf{Z}} \quad (15)$$

From Eq. (4) the elements of $\bar{\mathbf{Z}}$ are

$$\bar{z}_{ik} = \left(\frac{E_N}{N_0} \right)^{1/2} \exp(j\theta_k) \quad (16)$$

Under certain conditions it is possible to determine the d_m 's and λ_m , so that Eq. (14) can be bounded. For instance, if N and p are both large (this corresponds to large bandwidth or low-rate codes) and $N \gg p$, these eigenvalues are (Ref. 3)

$$\lambda_m = \frac{\sin 2\pi m \frac{N}{p}}{2\pi m \frac{N}{p}} \quad (17)$$

Also the k th element of the m th eigenvector is

$$e_{mk} = \frac{1}{(N)^{1/2}} \exp\left(j2\pi \frac{km}{N}\right) \quad (18)$$

for $0 \leq m \leq N/p$. Therefore, the elements of the m th eigenvector are complex exponentials with frequency m cycles per code word. With reference to Eq. (15), the d_m 's are the coefficients of the Fourier series expansion of $\bar{\mathbf{Z}}$.

As an example, consider the case where the channel phase varies linearly with time at the rate of m' cycles per code word; then

$$\bar{z}_{ik} = \left(\frac{E_N}{N_0} \right)^{1/2} \exp\left(j2\pi \frac{km'}{N}\right) \quad (19)$$

Since the elements of the mean vector $\bar{\mathbf{Z}}$ vary sinusoidally, there is only one term in the Fourier series expansion for $\bar{\mathbf{Z}}$; hence, from Eq. (15), there is only one non-zero d_m , that is

$$d_m = \left(\frac{NE_N}{N_0} \right)^{1/2}, \quad m = m' \quad (20)$$

$$= 0, \quad \text{otherwise}$$

Using this simplification, the two-signal error probability can be upper-bounded as (Ref. 3)

$$P_2(\epsilon) \leq \exp(-NR_0) \quad (21a)$$

where

$$R_0 = \max_{\substack{0 \leq \eta \leq 1 \\ 0 \leq \tau}} \left[\frac{E_N}{N_0} \frac{\eta \frac{\sin 2\pi f\tau}{2\pi f\tau}}{1 + \eta \frac{\sin 2\pi f\tau}{2\pi f\tau}} - \frac{1}{2\tau \left(\frac{S}{N_0} \right)} \ln \left(\frac{1}{1 - \eta^2} \right) \right] \quad (21b)$$

Here f is the rate of channel phase change in hertz and τ is the length of time over which each of the p -symbol phase references Q_k extends. For a given signal-to-noise ratio S/N_0 and rate of phase change f , there exists an optimum reference duration τ . The reason for this is as follows: for $\tau \ll 1/f$, the energy-to-noise ratio in the reference $\tau S/N_0$ is small; hence, it does not provide reliable estimates of true channel phase. On the other hand for large τ , although $\tau S/N_0$ is large, the channel phase changes radically in the duration of the references; hence, again its prediction of present phase is unreliable. These two effects can be observed separately in the two terms comprising R_0 in Eq. (21b). The first term represents the effect of the channel phase variation over the duration of the reference; it tends to decrease with increasing τ . The second term is the degradation due to noise in the phase estimate; it increases toward zero with increasing τ (increasing reference energy).

The two-signal error probability can be extended to $M = \exp(NR_N)$ signals by using the "union bound" (Ref. 1). The overall error probability is then

$$P(\epsilon) \leq MP_2(\epsilon) = \exp[-N(R_0 - R_N)] \quad (22)$$

where R_N is the code rate in nats/channel symbol. Error probability is therefore guaranteed to decrease exponentially with code length N for rates less than R_0 .

3. Sequential Decoding Bounds

Sequential decoding is a practical procedure for communicating with error probabilities which decrease exponentially with code length. Sequential decoding for discrete memoryless channels has been investigated extensively (Refs. 1, 4, 5, and SPS 37-32, Vol. IV, p. 303). Its utility lies in the fact that for rates less than a certain "computational cutoff rate," the average computational effort per information bit decoded is small and is independent of the code length.

For the channel with time varying phase, similar results can be obtained (Ref. 3). For instance, the probability of

undetectable error $P'(\epsilon)$ can be bounded in a form almost identical to the block code error probability

$$P'(\epsilon) \leq K \exp[-N(R_0 - R_N)] \quad (23)$$

where K is a constant. Of equal importance, the distribution of the number of decoder computations per bit decoded c is Pareto, that is

$$\Pr[c \geq L] \leq K' L^{-\alpha} \quad (24)$$

$$\alpha \approx \frac{R_0}{R_N}$$

For rates $R_N > R_0$, the average computation per bit decoded \bar{c} is unbounded. Therefore, R_0 can be interpreted as the maximum usable rate, or in the accepted terminology the computational cutoff rate. Reducing error probability to any desired level is no problem with sequential decoding because the code length N can be made quite large with little increase in equipment complexity. Therefore, instead of using error probability as a measure of system performance, it is more meaningful to use the energy per bit E_b required for reliable communication based on the computation problem. If R_0 is the highest usable rate

$$E_b = \frac{E_N}{R_0 \log_2 e} = \frac{\text{energy/symbol}}{\text{bits/symbol}} = \text{energy/bit} \quad (25)$$

A lower bound on E_b is the energy per bit required for a channel with known phase operating at channel capacity. This has been shown to be (Ref. 1) $E_{b,min} = N_0 \ln 2$. A comparison of E_b to $E_{b,min}$ will now be made for an interesting class of random phase processes.

4. Practical System Considerations

In a practical communication channel, such as the telemetry channel of a deep space communication link, the phase is, in general, a random process rather than a deterministic waveform. Suppose that $\theta(t)$ is such that $\cos[\omega_0 t - \theta(t)]$ is band-limited to frequencies between $2\pi\omega_0 - W_p$ and $2\pi\omega_0 + W_p$. This would be the case, for instance, if $r(t)$ were tracked by a squaring or Costas loop (Ref. 6); in that case $2W_p$ would correspond to the loop bandwidth. This might also be the case if the local oscillator in the receiver were controlled by an ephemeris rather than a phase-locked loop. At any rate, whenever $\cos[\omega_0 t - \theta(t)]$ is band-limited, the Fourier coefficients d_m of the mean vector \mathbf{Z} , are zero for m greater than some m' . Now, it is apparent from Eq. (21b) that in the linearly changing phase case, R_0 decreases monotonically

with increasing rate of phase change f . Therefore, since W_p is the highest rate of phase change possible in the band-limited case, R_0 for this case can be lower-bounded by that for a linearly changing phase of $f = W_p$ hertz. For the band-limited case, from Eq. (21b)

$$R_0 \geq \max_{\substack{0 \leq \eta \leq 1 \\ 0 \leq \tau}} \frac{E_N}{N_0} \left[\frac{\eta \frac{\sin 2\pi W_p \tau}{2\pi W_p \tau}}{1 + \eta \frac{\sin 2\pi W_p \tau}{2\pi W_p \tau}} - \frac{1}{2W_p \tau \left(\frac{S}{W_p N_0} \right)} \ln \left(\frac{1}{1 - \eta^2} \right) \right] \quad (26)$$

Operating at this rate, the energy bit required, from Eq. (25), is shown in Fig. 4 as a function of $S/2W_p N_0$, the signal-to-noise ratio in the bandwidth of the phase process. In this figure E_b is normalized by $E_{b,min} = N_0 \ln 2$.

The curve in Fig. 4 asymptotically approaches 3 dB as $S/2W_p N_0$ gets larger. Here, $E_b/E_{b,min} = 3$ dB is the limit imposed by a system operating with actually known phase at half of channel capacity [$R_0 = C/2$ for this case (Ref. 1)]. Thus, in Fig. 4, $E_b/(E_{b,min} - 3 \text{ dB})$ represents the *extra* energy per bit required due to imperfect knowledge of channel phase. For instance, if $S/2W_p N_0 = 18$ dB, about 1 dB is lost due to the random phase. Since the curve in Fig. 4 is based only on the lower bound on R_0 in Eq. (26), it represents an upper bound on $E_b/E_{b,min}$. Recall that this curve was derived using only the fact that $\cos \theta(t)$ is band-limited. The particular statistics of $\theta(t)$ were not accounted for; therefore, an actual $E_b/E_{b,min}$ curve may lie somewhere beneath that of Fig. 4. The important point is that Fig. 4 presents an upper bound valid for any band-limited process whatever. Note that

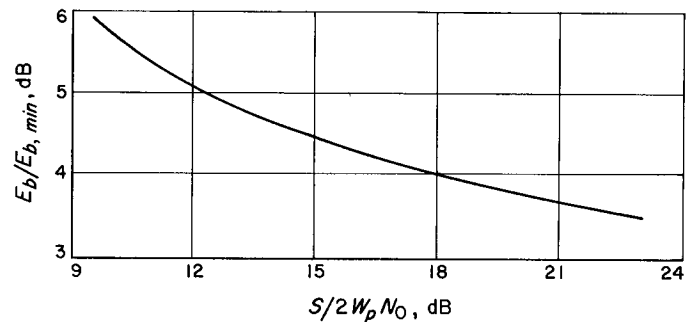


Fig. 4. Normalized energy per bit required as a function of the signal-to-noise ratio in the bandwidth of $\cos \theta(t)$

the analysis does not take into account degradation due to receiver output quantization, and assumes a low-code rate R_N in bits/channel symbol. The effects of these two factors on performance is expected to be comparable to that for memoryless channels (SPS 37-32, Vol. IV, p. 303).

5. Discussion

Several methods have previously been proposed for estimating channel phase from a biphas-modulated received signal. One method that does not make use of decision direction is the Costas loop or its equivalent (Refs. 6 and 7). This method has one serious drawback: system performance deteriorates as the bandwidth of the biphas modulation increases. For low rate codes, which are most efficient, there are many channel symbols per bit; the modulation bandwidth is, therefore, relatively large and performance is severely degraded.

A decision-directed scheme similar to the one described here was suggested for uncoded systems (Ref. 2). This scheme has the disadvantage that wrong decisions are never corrected. They degrade performance in a manner that varies with the error rate.

In a coded system decision direction does not suffer from these shortcomings. If the decoder is observing the correct code word (correct path in convolutional codes) all "decisions" used in forming the phase reference are correct. The modulation is, therefore, effectively removed and performance, as shown in Fig. 4, is independent of the symbol rate (modulation bandwidth) and information bit rate. It depends only on the ratio of signal power to noise power in the bandwidth of the phase process.

References

1. Wozencraft, J. M., and Jacobs, I. M., *Principles of Communication Engineering*. John Wiley & Sons, New York, 1965.
2. Proakis, J. G., et al., "Performance of Coherence Detection Systems Using Decision Directed Channel Measurement," *IEEE Trans. Commun. Technol.*, COM-5, March 1957.
3. Heller, J. A., *Sequential Decoding for Channels with Time Varying Phase*, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, Mass., Sep. 1967.
4. Savage, J. E., "The Distribution of the Sequential Decoding Computation Time," *IEEE Trans. Inform. Theory*, IT-12, April 1966.
5. Jordan, K. L., "The Performance of Sequential Decoding in Conjunction with Efficient Modulation," *IEEE Trans. Commun. Technol.*, COM-14, June 1966.
6. Viterbi, A. J., *Principles of Coherent Communication*. McGraw-Hill Book Co., Inc., New York, 1966.
7. Haccoun, D., "Simulated Communication with Sequential Decoding and Phase Estimation," M. S. thesis, Massachusetts Institute of Technology, Cambridge, Mass., Sep. 1966.

B. Synchronization of PCM Channels by the Method of Word Stuffing, S. Butman

1. Introduction

In digital communication it is often necessary to time-division multiplex a number of digital channels with different and fluctuating pulse rates into a single high-rate link, and to invert this operation (demultiplex or decommutate) at various receiving points down the line. In order for this process to function properly, it is necessary to have the pulses in each channel occur at specified instants in time so that they may be sampled in an order known to the receiving stations. This retiming process is called synchronization. It can be effected by reading the pulses of each channel into an elastic memory buffer (to be discussed later) and reading out the contents of each buffer in sequence.

From the viewpoint of an individual channel, synchronization is equivalent to rate equalization since the fluctuating buffer input rate $r(t)$ is converted to the synchronous buffer output rate s . Buffer overflow, which leads to loss of data, can be prevented by having $s \geq \max_t r(t)$. The difference rate

$$s - r(t) \geq 0 \quad (1)$$

is the buffer depletion rate, and it must be made up by inserting extra pulses (or time slots) into the output stream. These *stuffed* pulses must be identified and deleted at the receiving stations (after decommutation) in order to prevent decoding errors. The transmission of this identification information requires extra channel capacity.

Several synchronization schemes using pulse stuffing have been reported (Refs. 1 to 6)¹. In each case a pulse is inserted into the output stream as soon as the memory buffer is empty. The methods differ, however, in the manner of signaling the presence of the stuffed pulse. In added bit signaling (Refs. 5 and 6), the signaling channel is created by adding an extra pulse for every N data pulses; this corresponds to an increment of $1/N$ in the total channel capacity. In statistical subcarrier signaling,¹ each of the I most probable data words is assigned two symbols A_i and B_i , $i = 1, 2, \dots, I$, and the presence of a stuffed pulse is signalled by substituting A_i 's for B_i 's.

¹Pan, J. W., U. S. patent pending.

In this case the amount of additional channel capacity set aside for signaling stuffed pulses is not clearly defined; however, it is evident that the introduction of extra symbols into the channel alphabet is equivalent to increasing the capacity.

The purpose of the present discussion is to generalize the concept of pulse stuffing to word stuffing (a word being defined as a sequence of k pulses) in order to reduce the capacity of the signaling channel. This requires that the buffer memory increases proportionately to the length of the stuffed word. The question of whether to use more memory and less signaling capacity or vice versa is a matter of economics of the particular situation.

2. Synchronization Using Stuffed Words

The operation of an elastic memory buffer is analogous to a reservoir which can be filled intermittently while being drained at a steady rate. Consider a buffer with a memory size of k pulses. When the contents of this buffer have been depleted to some reference, say zero, the buffer must be allowed to refill by a temporary pause in its read-out. At the same time, a predetermined sequence of k pulses, a word W , is inserted into the output stream in order to maintain the synchronous rate s . (Actually, there is no such thing as not inserting pulses, since the absence of a signal is part of the code.) The time interval involved in the word stuffing operation is k/s , and during this time interval the input stream supplies

$$\begin{aligned} k' &= \int_0^{k/s} r(t) dt \\ &= k - \int_0^{k/s} [s - r(t)] dt \\ &\leq k \end{aligned} \quad (2)$$

pulses for reading into the buffer. Because of the discrete nature of read-in and read-out mechanisms, k' is either the greatest integer in, or the nearest integer to the quantity on the right-hand side of Eq. (2). If r is the average value of $r(t)$, then the average value of k' is kr/s .

The average rate of occurrence of stuffed words is equal to the average depletion rate divided by the average number of pulses read in, that is, $s(s-r)/kr$. Thus, although the exact time of occurrence of a stuffed word is unpredictable, stuffed words occur on the average only one k th as often as stuffed pulses, and require only one k th the signaling rate.

An even better approach is to signal the occurrence of a data word W rather than the stuffed word. Thus, in the absence of signaling, the receiver is arranged to detect and delete the sequence W , and signaling is used to inhibit deletion when there is a data sequence W . In this way, only the stuffed word W is removed, provided that there are no transmission errors. For a non-zero channel error probability, the situation becomes somewhat more complicated due to the possibility of errors of the first and second kind (non-detection of W and false alarms); this matter will be taken up in Part 3, below. If the data stream consists of m equiprobable pulses, the probability of a data word being W is m^{-k} and the average signaling rate is only $r/(km^k)$, which is smaller than the rate of stuffed words $s(s-r)/kr$ when $k > \log_m [r^2/s(s-r)]$. Finally, it is of interest to observe that signaling may be discarded at the price of allowing the error probability to increase by about m^{-k} . This is accomplished by forcing the transmitter to change one (or more) pulses in the data sequence W . If the channel were error-free, all data would pass through the W detector at the receiver; however, the data word that was W will have one pulse error, causing the average error probability to increase by m^{-k}/k . The number of intentional errors will increase when redundancy is introduced to combat errors of the first and second kind.

3. Errors of the First and Second Kind

Let p be the error probability for a binary symmetric channel ($m = 2$). The probability of not detecting the stuffed word is then

$$P_1 = 1 - (1 - p)^k \quad (3)$$

and it increases with k . It can, however, be decreased by redundant coding of W . Thus, the receiver is instructed to take any sequence of k pulses with a Hamming distance d or less from W as the stuffed word itself. This, of course, increases the number of data words that will be mistaken for W , unless they are intentionally put in error at the transmitter, or unless the signaling rate is increased. The number of errors that the transmitter must introduce to change the distance from i to $d + 1$ is $d + 1 - i$, and the number of words at a distance i is $\binom{k}{i}$. Therefore, the intentional average probability of error is

$$P_1 = \frac{2^{-k}}{k} \left[\sum_{i=0}^d (d + 1 - i) \binom{k}{i} \right] \quad (4)$$

or the average rate of the inhibit signal is rP_I . However, the probability of an error of the first kind is now only

$$P_1 = \sum_{i=d+1}^k \binom{k}{i} p^i (1-p)^{k-i} \quad (5)$$

A false alarm, or an error of the second kind, is the event that a sequence of k data pulses originally at a distance $h \geq d+1$ is in the course of transmission altered to a sequence whose distance from the stuffed sequence is $\leq d$. The probability of error of the second kind P_2 increases with d , its minimum value being equal to P_1 when d is zero. In order to calculate P_2 , note that all but

h of the pulses in a sequence at distance h from W are the same as in W . Therefore, the distance decreases by i when there are i errors in the h unlike pulses, and increases by j when there are j errors in the remaining $k-h$ pulses. Since there are $\binom{h}{i}$ ways of making i errors in the h unlike pulses, and $\binom{k-h}{j}$ ways in the remaining $k-h$ pulses, the probability of changing the distance from h to $h+j-i$ is

$$p \{h \rightarrow h+j-i\} = \binom{h}{i} \binom{k-h}{j} p^{i+j} (1-p)^{k-i-j} \quad (6)$$

and the probability of having $j-i \leq d-h$ is

$$p \{j-i \leq d-h\} = \sum_{i=h-d}^h \sum_{j=0}^{i+d-h} \binom{h}{i} \binom{k-h}{j} p^{i+j} (1-p)^{k-i-j} \quad (7)$$

Finally, since the number of words at distance h from W is $\binom{k}{h}$ except that for $h = d+1$ there are an additional

$$\sum_{h=0}^d \binom{k}{h}$$

words that were originally at a distance $\leq d$, the probability of a false alarm is

$$P_2 = 2^{-k} \left[p \{j-i \leq 1\} \sum_{h=0}^{d+1} \binom{k}{h} + \sum_{h=d+2}^k \binom{k}{h} p \{j-i \leq d-h\} \right] \quad (8)$$

4. Performance

The total increment in the average error probability is then

$$\Delta p = \frac{s(s-r)P_1}{r^2} + P_2 + P_I \quad (9)$$

For a given value of k , the above expression will have a minimum with respect to d because P_1 decreases with d while P_2 and P_I increase. In many practical situations $\Delta p \approx P_I$. For example, consider a channel with

$$0 \leq (s-r)/r \leq 1$$

and an error probability $p = 10^{-6}$. Then, if $k = 30$ and $d = 1$, P_1 and P_2 are negligible compared to P_I which is $\sim 2^{-30} \approx 10^{-9}$, so that Δp is only 0.1% of p ; and this is with no signaling.

It might be argued that errors of the first or second kind are the most catastrophic, since their occurrence upsets the data flow, causing, possibly, serious difficulties in decoding. It is, therefore, of interest to minimize the average rate of occurrence of undetected stuffed words or false alarms. The relevant expression, when $s \approx r$, so that $s(s-r)/r^2 \approx (s-r)/r$, is

$$u = \frac{[(s-r)P_1 + rP_2]}{k} \quad (10)$$

where u is the average rate of occurrence of errors of the first or second kind. Table 1 presents u in events/year, for a channel with $p = 10^{-6}$, $s = 1.544 \times 10^6$ pps, and $s-r = 200$ pps as a function of k and d . The minimum value of u for each choice of k is underlined, indicating the best choice for d .

Table 1. Yearly occurrence of errors of the first or second kind as a function of k and d for a channel with $P = 10^{-6}$, $s = 1.544 \times 10^6$ pps, and $s - r = 200$ pps

$d \backslash k$	10	20	30	40
0	60,500	6,500	6,500	6,500
1	5,500,000	1,100	1.4	1.3
2		5,500	20	0.04
3				4.0

5. Conclusion

This article investigated the performance of rate equalization schemes based on the concept of pulse stuffing. It was shown that by increasing the length of the sequence of stuffed pulses, a previously neglected quantity, it is possible to reduce, and even to eliminate, the need for an auxiliary signaling channel. The analysis presented for the binary symmetric channel is a worst-case solution because each data sequence was assumed to be equally likely. If non-uniform statistics prevail, the stuffed sequence should be the least probable data word. Finally, it should be noted that word stuffing is particularly suitable in equalizing channels with large fluctuations in rate.

References

1. Mayo, J. S., "PCM Network Synchronization," U. S. Patent 3136861, 1964.
2. Mayo, J. S., "Experimental 224 Mb/s PCM Terminals," *Bell Sys. Tech. J.*, Vol. 44, pp. 1813-1841, Nov. 1965.
3. Witt, F. J., "An Experimental 224 Mb/s Digital Multiplexer-Demultiplexer Using Pulse Stuffing Synchronization," *Bell Sys. Tech. J.*, Vol. 44, pp. 1843-1885, Nov. 1965.
4. Geigel, A. A., and Witt, F. J., "Elastic Stores in High Speed Digital Systems," *NEREM Rec.*, Vol. 6, 1964.
5. Johannes, V. I., "Multiplexing of Asynchronous Digital Signals Using Pulse Stuffing With Added-Bit Signaling," *IEEE Trans. Commun. Technol.*, Vol. COM-14, No. 5, pp. 562-568, Oct. 1966.
6. Mayo, J. S., "An Approach to Digital System Networks," *IEEE Trans. Commun. Technol.*, Vol. COM-15, No. 2, pp. 307-310, Apr. 1967.

C. Factoring Polynomials Over Finite Fields,

R. J. McEliece

1. Introduction

In this article an algorithm for factoring polynomials over finite fields will be given. Of course, the existence

of such an algorithm is not in doubt, since it is clearly possible to recursively generate all irreducible polynomials of a given degree over a given finite field. However, the algorithm given here is quite practical, and does not require a table of irreducible polynomials. The algorithm is very well suited for calculating the factors of $x^n + 1$ over $GF[2]$; these factors are the characteristic polynomials for the linear recurrences used to generate all (n, k) cyclic codes, and a table of these factors for $n \leq 100$ is included.

2. The Algorithm

Throughout, let $F = GF[q]$, $q = p^r$, p a prime. If $f(x)$ and $g(x)$ are two polynomials over F , denote by (f, g) their greatest common divisor, which will be assumed to be monic. We are given a polynomial $f(x)$, and asked to write f as a product of irreducible factors; we are free to assume that $f(x)$ is squarefree, since unless f is a p th power $f/(f, f')$ will be a nontrivial squarefree divisor of f . Let us further assume that $f(0) \neq 0$. Under these circumstances, there will be a least integer e , such that $f(x) | x^e - 1$. Since f is squarefree, $p \nmid e$; e is called the *exponent* of f .

Definition. For each i , let n_i be the least integer such that $x^i \equiv x^{iq^{n_i}} \pmod{f(x)}$. If the exponent e is known, the n_i are given by $n_i = \text{ord}_{e/(e, i)}(q)$, but, of course, it is not necessary to know e in order to compute the n_i . We define "test polynomials" for f as follows:

$$T_f^{(i)}(x) \equiv x^i + x^{iq} + \cdots + x^{iq^{n_i-1}} \pmod{f(x)}$$

Next, if \bar{n}_i is the least integer such that

$$x^i \equiv x^{iq^{\bar{n}_i}} \pmod{x^e - 1}$$

define

$$T_e^{(i)}(x) \equiv x^i + x^{iq} + \cdots + x^{iq^{\bar{n}_i-1}} \pmod{x^e - 1}$$

(We emphasize that f always represents a polynomial, e an integer.) Then, since $n_i | \bar{n}_i$, for suitable integers m_i , it will be true that

$$T_e^{(i)}(x) \equiv m_i T_f^{(i)}(x) \pmod{f(x)} \quad (1)$$

Theorem 1 is the heart of the algorithm.

Theorem 1. If $h(x)^q \equiv h(x) \pmod{f(x)}$, then

$$f(x) = \prod_{a \in F} (f(x), h(x) - a)$$

Proof. Let θ be a root of f in a splitting field K . Then $h(\theta)^q = h(\theta)$ and so $h(\theta)$, being fixed by the Galois group of K/F , is an element of F . Thus, every root of f is a root of exactly one of the polynomials $h(x) - a$, and theorem 2 follows.

Corollary.

$$f(x) = \prod_{a \in F} (f(x), T_f^{(i)}(x) - a)$$

The final result needed is that the factorizations provided by the corollary are sufficient to separate all the irreducible factors of f . Several easy preliminary results are needed.

Lemma 1. If $h(x)^q \equiv h(x) \pmod{x^e - 1}$, then $h(x)$ is an F -linear combination of the polynomials $T_e^{(i)}(x) \pmod{x^e - 1}$.

Proof. Suppose

$$h(x) \equiv \sum_{k=0}^{e-1} b_k x^k \pmod{x^e - 1}$$

Then,

$$h(x)^q \equiv h(x^q) \equiv \sum_{k=0}^{e-1} b_k x^{qk} \pmod{f(x)}$$

Hence, if $k_1 \equiv k_2 q^t \pmod{e}$ for some $t > 0$, $b_{k_1} = b_{k_2}$. Hence, $h(x)$ is an F -linear combination of the $T_e^{(i)}(x)$, as asserted.

Lemma 2. If f_1 is an irreducible divisor of $x^e - 1$, then there is a polynomial $g(x)$ such that $(f_1 g)^q \equiv (f_1 g) \pmod{x^e - 1}$, and $(f_1 g, x^e - 1) = f_1$.

Proof. Since $x^e - 1$ is squarefree, then

$$(f_1, (x^e - 1)/f_1) = 1$$

and so a polynomial $g(x)$ may be found such that $f_1 g \equiv 1 \pmod{(x^e - 1)/f_1}$. Then $(f_1 g)^2 \equiv f_1 g \pmod{x^e - 1}$ and so also $(f_1 g)^q \equiv f_1 g \pmod{x^e - 1}$. Finally, since $(g, (x^e - 1)/f_1) = 1$, it follows that $(f_1 g, x^e - 1) = f_1$, completing the proof of the lemma. We are now ready to prove the "separation" theorem.

Theorem 2. Let f_1 and f_2 be distinct irreducible divisors of f . Then there is an integer i and distinct elements a, b of F such that

$$f_1 | (T_f^{(i)} - a); \quad f_2 | (T_f^{(i)} - b)$$

Proof. Suppose by way of contradiction that for each i there is an element $a_i \in F$ such that $f_1 f_2 | T_f^{(i)} - a_i$. By lemma 2 there exists a polynomial $h(x)$ such that

$$(f_1 h)^q \equiv f_1 h \pmod{x^e - 1}$$

and $(f_1 h, x^e - 1) = f_1$. Thus, by lemma 1,

$$f_1 h \equiv \sum b_i T_e^{(i)}(x) \pmod{x^e - 1}$$

But by Eq. (1)

$$f_1 h \equiv \sum m_i b_i T_f^{(i)}(x) \pmod{f(x)}$$

and by assumption

$$f_1 f_2 | \sum m_i b_i (T_f^{(i)} - a) \equiv f_1 h - b \pmod{f(x)}$$

where

$$b = \sum m_i b_i a_i \in F$$

But this is in conflict with $(f_1 h, f) = f_1$, and so completes the proof.

Theorems 1 and 2 allow us to factor any polynomial $f(x)$ by the following steps:

F1. Eliminate powers of x .

F2. Check to see whether or not f is a perfect p th power; if it is, reduce it.

F3. Compute $f/(f, f')$ and apply the following steps to it.

F4. Compute the $T_f^{(i)}(x)$.

F5. Find one non-trivial factorization provided by the corollary to theorem 1. (If all the $T_f^{(i)}$ are elements of F , f is irreducible.)

F6. Reduce the $T_f^{(i)}$ modulo the factors of f found at F5 and apply theorem 1 again. Continue until all the irreducible factors of f are found.

F7. Find the highest power of the irreducible factors found at F5 and F6 which divide the original f . Apply F2 to what remains.

3. An Example

Let us apply the algorithm to the polynomial

$$f(x) = x^{17} + x^{14} + x^{13} + x^{12} + x^{11} + x^{10} + x^9 + x^8 + x^7 + x^5 + x^4 + x + 1$$

over $GF[2]$: $f(0) = 1$ so we proceed to $F2$, and find

$$f' = x^{16} + x^{12} + x^{10} + x^8 + x^6 + x^4 + 1$$

To compute (f, f') , we use Euclid's algorithm: we abbreviate a polynomial

$$\sum_{i=0}^n a_i x^i$$

by $(a_n a_{n-1} \cdots a_1 a_0)$

$$\begin{array}{r} 100111111110110011 \\ 10001010101010001 \\ \hline 101010100010001 \\ 10001010101010001 \\ \hline 100000100010101 \\ 101010100010001 \\ \hline 1010000000100 \\ 101010100010001 \\ \hline 10100000001 \\ 10100000001 \end{array}$$

Hence, $(f, f') = x^{10} + x^8 + 1$, and an easy division gives

$$\frac{f}{(f, f')} = x^7 + x^5 + x^4 + x + 1 = \bar{f}$$

which we now know to be squarefree. To compute successive squares, it is convenient to have a list of even powers of x modulo $\bar{f}(x)$:

$$\begin{array}{ll} x^0 & 0000001 \\ x^2 & 0000100 \\ x^4 & 0010000 \\ x^6 & 1000000 \\ x^8 & 1100110 \\ x^{10} & 1001101 \\ x^{12} & 1010010 \\ \\ x & 0000010 \\ x^2 & 0000100 \\ x^4 & 0010000 \end{array}$$

$$\begin{array}{ll} x^8 & 1100110 \\ x^{16} & 0001011 \\ x^{32} & 1000101 \\ x^{64} & 1000011 \\ x^{128} & 1010111 \\ x^{256} & 0100001 \\ x^{512} & 1001100 \\ & \hline & 1000111 \end{array}$$

$$(x^{210} = x)$$

$$\begin{array}{ll} x^3 & 0001000 \\ x^6 & 1000000 \\ x^{12} & 1010010 \\ x^{24} & 0110000 \\ x^{48} & 0101011 \\ & \hline & 0000001 \end{array}$$

$$(x^{3 \cdot 2^8} = x^3)$$

$$\begin{array}{ll} x^5 & 0100000 \\ x^{5 \cdot 2} & 1001101 \\ x^{5 \cdot 2^2} & 0000011 \\ x^{5 \cdot 2^3} & 0000101 \\ x^{5 \cdot 2^4} & 0010001 \\ x^{5 \cdot 2^5} & 1100111 \\ x^{5 \cdot 2^6} & 0001010 \\ x^{5 \cdot 2^7} & 1000100 \\ x^{5 \cdot 2^8} & 1000110 \\ x^{5 \cdot 2^9} & 1010110 \\ & \hline & 1000111 \end{array}$$

$$(x^{5 \cdot 2^{10}} = x^5)$$

Hence,

$$T^{(1)}(x) = T^{(5)}(x) = x^6 + x^2 + x + 1$$

and

$$\begin{aligned} \bar{f}(x) &= (x^7 + x^5 + x^4 + x + 1, x^6 + x^2 + x + 1) \\ &\times (x^7 + x^2 + x + 1, x^5 + x + 1) \end{aligned}$$

must be a factorization into irreducibles:

$$\begin{array}{r}
 10110011 \\
 1000111 \\
 \hline
 111101 \\
 1000111 \\
 \hline
 111101 \\
 \boxed{1111011} \\
 \\
 10110011 \\
 1000111 \\
 \hline
 111111 \\
 100011 \\
 \hline
 11100 \\
 100011 \\
 \hline
 11011 \\
 11011 \\
 \hline
 111 \\
 \boxed{111}
 \end{array}$$

Hence,

$$\bar{f}(x) = (x^5 + x^4 + x^3 + x^2 + 1)(x^2 + x + 1)$$

as a product of irreducibles. Next we check to see whether or not (f, f') is divisible by either of the two factors already found.

$$(f, f') = (x^5 + x^4 + 1)^2$$

so that we need only try $(x^2 + x + 1)^2$. It is easily seen that

$$x^5 + x^4 + 1 = (x^2 + x + 1)(x^3 + x + 1)$$

Hence,

$$f(x) = (x^5 + x^4 + x^3 + x^2 + 1)(x^3 + x + 1)^2(x^2 + x + 1)^3$$

is the required factorization.

4. Factoring $x^n + 1$ Modulo 2

As a less trivial example of the algorithm, consider the factorization of the polynomials $x^n + 1$ over $GF[2]$. In this case the computation of the test polynomials is very

Table 2. Polynomials of Period n over $GF[2]$

n	Factors
3	irreducible
5	irreducible
7	13
9	$(3 \cdot 3)$
11	irreducible
13	irreducible
15	31
17P	471,727
19	irreducible
21	165
23	5343
25	$(5 \cdot 5)$
27	$(3 \cdot 9)$
29	irreducible
31	75,73,45
33P	3043,2251
35	16475
37	irreducible
39	17075
41P	5747175,6647133
43P	64213,47771,52225
45	$(15 \cdot 3)$
47	43073357
49	$(7 \cdot 7)$
51	637,661
53	irreducible
55	7164555
57P	1735357,1341035
59	irreducible
61	irreducible
63	147,141,155
65P	15353,13535,12345,10761
67	irreducible
69	34603145
71	503700420663
73	1401,1641,1511,1145
75	$(15 \cdot 5)$
77	16471647235
79	11435717264067
81	$(3 \cdot 27)$
83	irreducible
85	771,613,735,675
87	3706175715
89	6061,7773,7571,7311
91	14015,15713,11721
93	3205,3247,2065
95	1435137342601
97P	10265044102212641, 17441554343330237
99	$(33 \cdot 3)$

simple; one needs only to compute the orbits of the integers modulo n under the permutations generated by $i \rightarrow 2i \pmod{n}$; these orbits contain the exponents which occur in the various test polynomials. For example with $n = 7$, the orbits are

$$(0), \quad (1, 2, 4), \quad (3, 6, 5)$$

and so the test polynomials are $x + x^2 + x^4$ and $x^3 + x^5 + x^6$. Using this method on an SDS-930 computer, it has been possible to obtain the complete factorization of $x^n + 1$ for $n \leq 350$; however, only the factorization for $n < 100$ is given in Table 2, and only the factors of $x^n + 1$ which are not factors of $x^m + 1$ for some $m < n$ are listed.

Polynomials are given the customary octal representation; e.g., 7053 represents

$$x^{11} + x^{10} + x^9 + x^5 + x^3 + x + 1$$

If a polynomial $f(x)$ divides $x^n + 1$, then so does its reciprocal polynomial, and only one member of a reciprocal pair is listed. For certain values of n , each polynomial is self-reciprocal; this is indicated by a "P" following the integer. Finally, for some values of n the factorization $x^{n \cdot m} + 1$ may be obtained from that of $x^n + 1$ by replacing x by x^m . This is indicated in Table 2 by the entry $(n \cdot m)$.

D. On Automorphism Groups of Block Designs,

R. E. Block²

1. Introduction

The theory of block designs has been studied both at JPL and elsewhere in connection with the construction of ranging sequences and orthogonal codes (SPS 37-25, Vol. IV, pp. 158-160; SPS 37-28, Vol. IV, pp. 232-234). The results of this article facilitate the search for such sequences and codes with desirable properties that allow rapid acquisition or decoding.

First, the key concept of a BIBD will be reviewed. A configuration D of "points" and "blocks" is called a BIBD with parameters v, b, r, k , and λ provided D has v points and b blocks, each point is on exactly r blocks, each block contains exactly k points, and each pair of distinct points occurs together in exactly λ blocks. An *automorphism* (also called *collineation*) of D is a pair of

permutations, one of the points and one of the blocks, preserving the incidence relations. The known direct methods for constructing BIBD's generally involve a group of automorphisms, e.g., the method of mixed differences (Ref. 1, Chap. 15 and Appendix).

A basic result on BIBD's (Fisher's inequality) says that $v \leq b$ unless the design is degenerate, i.e., unless each block contains all v points, or, equivalently, $r = \lambda$ (assuming that $v > 1$ and $b > 0$). Suppose that there is a group G of automorphisms of D , with t point orbits, of lengths v_1, \dots, v_t , and t' block orbits, of lengths $b_1, \dots, b_{t'}$. Fisher's inequality may be regarded as an assertion about the orbits for the identity group. One generalization of this, which we proved in Ref. 2, Corollary 2.2, says that $t \leq t'$ if D is nondegenerate. In the present article we shall give a generalization which involves not only the number of orbits but also their lengths.

If D is *symmetric* ($v = b$) then $t = t'$ (Refs. 3, 4, and 5) and then the orbit decomposition is also called *symmetric* if, after suitable reordering, $v_i = b_i$ for $i = 1, \dots, t$. Sufficient conditions for this are that G be a cyclic group (Ref. 5), or else a p -group where $p \nmid r - \lambda$ (Ref. 2). More generally, suppose that p is a prime, and denote by v_p the (exponential) p -adic valuation (say of the integers Z), so that if $a \in Z$ then

$$a = p^{v_p(a)} a'$$

where $(p, a') = 1$. We proved in Ref. 2 for symmetric designs that if $p \nmid r - \lambda$ then, after reordering, $v_p(v_i) = v_p(b_i)$ for $i = 1, \dots, t$.

Our main result in the present article applies to not necessarily symmetric designs and states: if $p \nmid (r - \lambda)(r, b) \times (k, v)$ then there is a reordering of the orbits so that $v_p(v_i) = v_p(b_i)$ (and so for a p -group $v_i = b_i$) for $i = 1, \dots, t$. The exclusion in the hypothesis of the cases $r \equiv \lambda \pmod{p}$, $r \equiv b \equiv 0 \pmod{p}$, and $k \equiv v \equiv 0 \pmod{p}$ is analogous to the exclusion in Fisher's inequality of the degenerate cases $r = \lambda$, $v = 1$, and $b = 0$ (and so $r = 0$), and $b > 0$, $v = 0$ (and so $k = 0$), respectively. We shall also prove: if $p \mid (r, b)(k, v)$ (in fact if just $p \mid rk$) but $p \nmid (r - \lambda)$ then there is a reordering so that $v_p(v_t) = 0$ and $v_p(v_i) = v_p(b_i)$ for $i = 1, \dots, t - 1$. Numerous examples (Ref. 1, Appendix) show that none of the restrictions on p can be removed.

We also note that a sharper form of Fisher's inequality is valid; there is a one-one mapping of the points to the blocks such that each point is incident with its image; we

²Consultant, Department of Mathematics, University of Illinois, Urbana, Ill.

shall also prove a similar property for the correspondences above of point orbits to block orbits.

We shall see that these results hold not only for the orbits of an automorphism group, but also for the point and block classes of a so-called *tactical decomposition* on the design. We shall also obtain similar results for other structures besides BIBD's, including the constant-distance codes and matrices, especially orthogonal codes and Hadamard matrices. Finally, we shall also apply our theorem on BIBD's to obtain results on permutation groups and on symmetric BIBD's, including projective planes. For a group G of permutations on a finite set Ω , Polya's enumeration theorem (Ref. 6) gives the number of orbits of the induced group G_k of permutations on the k -element subsets of Ω in terms of the cycle structures of the elements of G , and our main theorem, applied to the (trivial) design of all these k -element subsets, will give information about the lengths of some of the orbits of G_k . For the case of a group of automorphisms of a projective plane, our theorem, applied to the designs of "flags" and "anti-flags" which we shall introduce, will give information about the lengths of the orbits of incident and non-incident point-line pairs.

2. The Main Theorem

Let $M = (m_{ij})$ be a $v \times b$ matrix with entries in a field F . Suppose that the set of row indices is the disjoint union of t nonempty subsets R_1, \dots, R_t , and that the set of column indices is the disjoint union of t' nonempty subsets $C_1, \dots, C_{t'}$. Then M is said to have a *tactical decomposition* (Ref. 2) with row classes R_i and column classes C_j , if for every i, j ($i = 1, \dots, t$; $j = 1, \dots, t'$) the submatrix (m_{hi}) ($h \in R_i, \ell \in C_j$) has constant column sums s_{ij} (right tactical decomposition) and constant row sums a_{ij} (left tactical decomposition). The $t \times t'$ matrix $S = (s_{ij})$ is called the *associated matrix of column sums*.

Then for a (generalized) incidence structure (i.e., finite set of points and blocks with an incidence relation between points and blocks), a *tactical decomposition* is just a partition of the points into point classes and of the blocks into block classes giving a tactical decomposition of the incidence matrix. For any group of automorphisms of an incidence structure, the point orbits are the point classes and the block orbits are the block classes of a tactical decomposition.

For any tactical decomposition, the number of elements in a row (point) class or in a column (block) class is called the *length* of this class; and the length of the row class R_i

will be denoted by v_i , and the length of C_j will be denoted by b_j .

We shall consider a non-archimedean valuation v on F , written exponentially, so that $v(\gamma\delta) = v(\gamma) + v(\delta)$ and $v(\gamma + \delta) \geq \min\{v(\gamma), v(\delta)\}$, $\gamma, \delta \in F$. For the application to block designs, F will be the rationals Q and v the p -adic valuation v_p , except for one case when $p = 2$. In the following lemma for any integer, m, I_m , and J_m will denote respectively the $m \times m$ identity matrix and the $m \times m$ matrix with all entries 1.

Lemma 1. Let F be a field with a non-archimedean valuation v , and let M be a $v \times b$ matrix over F with entries in the valuation ring at v . Suppose that M has a tactical decomposition with row class lengths v_1, \dots, v_t and column class lengths $b_1, \dots, b_{t'}$, where $t \leq t'$, and that there are elements α and β in F such that $MM' = \alpha I_v + \beta J_v$. Then there is a reordering of the column classes such that

$$\sum_{i=1}^t |v(b_i) - v(v_i)| \leq v((\alpha + v\beta) \alpha^{t-1}) \quad (1)$$

Proof. Counting in two ways the sum of the entries of the submatrix (m_{hi}) ($h \in R_i, \ell \in C_j$) of M , we have

$$v_i a_{ij} = s_{ij} b_j$$

and hence

$$v(v_i) + v(a_{ij}) = v(s_{ij}) + v(b_j) \quad (2)$$

As in lemma 5.1 of SPS 37-25, Vol. IV, pp. 158-160, we have the fundamental matrix equation

$$SBS' = \alpha VI_t + \beta VJ_t V$$

where B and V denote, respectively, the $t' \times t'$ and $t \times t$ diagonal matrices $\text{diag}(b_1, \dots, b_{t'})$ and $\text{diag}(v_1, \dots, v_t)$. Using the Binet-Cauchy formula for the determinant of the product of a $t \times m$ and an $m \times t$ matrix, and the determination of $\det(\alpha VT_t + \beta VJ_t V)$ as in Ref. 2, p. 40, we have

$$\sum_{1 \leq j_1 < \dots < j_t \leq t'} b_{j_1} \dots b_{j_t} (\det S(j_1, \dots, j_t))^2 = v_1 \dots v_t (\alpha + v\beta) \alpha^{t-1}$$

where $S(j_1, \dots, j_t)$ denotes the submatrix of S obtained by deleting all but the t columns indicated.

For notational convenience we reorder the column classes so that one of the summands of the left side having minimal value of v is the one for which $(j_1, \dots, j_t) = (1, \dots, t)$. We also write

$$\begin{aligned} v(v_1, \dots, v_t) &= c_v, & v(b_1, \dots, b_t) &= c_b \\ v(\det S(1, \dots, t)) &= c_d, & v((\alpha + v\beta)\alpha^{t-1}) &= c_\alpha \end{aligned} \quad (3)$$

$$c_b + 2c_d \leq c_v + c_\alpha$$

There is some transversal $s_1, \pi(1), s_2, \pi(2), \dots, s_t, \pi(t)$ of $S(1, \dots, t)$ such that

$$v\left(\prod_{i=1}^t s_{i, \pi(i)}\right) \leq c_d$$

and by reordering the columns of S we can suppose that π is the identity permutation. Now write

$$c_1 = \sum_{i=1}^t \max\{v(b_i) - v(v_i), 0\}$$

$$c_2 = \sum_{i=1}^t \max\{v(v_i) - v(b_i), 0\}$$

Then $c_b = c_v + \sum (v(b_i) - v(v_i)) = c_v + c_1 - c_2$, and hence $c_v + c_1 - c_2 + 2c_d \leq c_v + c_\alpha$. But by Eq. (2), $c_2 \leq v(s_{11} \dots s_{tt}) \leq c_d$, and so $c_1 \leq c_\alpha - c_d$ and $c_1 + c_2 \leq c_\alpha - c_d + c_d$, which gives Eq. (1). We can now state the main theorem.

Theorem 1. Let D be a BIBD with parameters v, b, r, k, λ , and let p be a prime not dividing $r - \lambda$. Suppose that D has a tactical decomposition. If $p \nmid (r, b)(k, v)$ then t distinct block classes C_{j_1}, \dots, C_{j_t} can be chosen such that the following conditions are satisfied for $i = 1, \dots, t$:

$$v_p(v_i) = v_p(b_{j_i}) \quad (4)$$

and, if $p \nmid rk$ or if $p \mid v_i$ then

$$s_i a_i \not\equiv 0 \pmod{p} \quad (5)$$

where s_i (resp. a_i) denotes the number of points (blocks) of R_i (C_{j_i}) incident with each block (point) of C_{j_i} (R_i). If $p \mid rk$, then for some ℓ , $p \nmid \ell$, and the t block classes can be chosen so that Eqs. (4) and (5) are satisfied for $i \neq \ell$.

The proof is omitted.

This theorem generalizes some results of Ref. 3; whereas the equalities of Eq. (4) are much sharper than the sets of inequalities of Ref. 2, pp. 47-48, for non-symmetric BIBD's, the present results hold for tactical decompositions, while the earlier results were also valid for *right* tactical decompositions.

Motivated by theorem 1, we define a *p-symmetry* of a tactical decomposition to be a one-one mapping of the set of row classes to the set of column classes such that Eq. (4) holds for $i = 1, \dots, t$ and we define a *p-semi-symmetry* to be such a mapping of the set of all but one, say R_ℓ , of the row classes such that Eq. (4) holds for $i \neq \ell$. A *p-symmetry* (resp. *p-semi-symmetry*) on a BIBD will be called *strong* if Eq. (5) holds for all i (resp. for all $i \neq \ell$).

While theorem 1 gives sharp information on the p -component of the lengths of t or $t-1$ of the block classes if $p \mid r - \lambda$, we can also give some information for every block (and point) class.

Proposition 1. Suppose there is given a tactical decomposition on a BIBD. Then for every block class C_j (resp. point class R_i) there is a point class R_i (resp. block class C_j) and integers $s (=s_{ij})$ and $a (=a_{ij})$ such that

$$v_i a = b_j s, \quad 0 < a \leq r, 0 < s \leq k$$

and in particular, for any prime q ,

$$v_q(b_j) = v_q(v_i) \text{ if } q > r, \quad v_q(b_j) \geq v_q(v_i) \text{ if } q > k$$

Proof. Take the tactical decomposition of the incidence matrix written with 1, 0. Given a $C_j, \sum_i s_{ij} = k$, and some $s_{ij} \neq 0$. Similarly given an $R_i, \sum_j a_{ij} = r$, and some $a_{ij} \neq 0$. The result then follows from Eq. (2).

Corollary 1. For the *p-symmetry* (resp. *p-semi-symmetry*) of theorem 1, the conclusions of proposition 1 hold for the pair R_i, C_{j_i} if $p \nmid rk$ or $p \mid v_i$ (resp. if $i \neq \ell$).

Proof. All that is needed is that $s_i \neq 0$, and this follows from Eq. (5).

Another application of lemma 1 is to Hadamard matrices, and more generally to constant-distance matrices, these being matrices with entries chosen from two symbols such that any two rows differ in the same number d of columns, where $d > 0$.

Theorem 2. Suppose that M is a $v \times b$ constant-distance matrix with distance d , and that M has a tactical decomposition. If $p \nmid 2d[2d + v(b - 2d)]$ (resp. $p \nmid 2d$) then the decomposition has a p -symmetry (resp. p -semi-symmetry). In particular, if M is Hadamard matrix of order v and if $p \nmid v$ then the decomposition is p -symmetric.

The proof is omitted.

3. Applications

Suppose that G is a group of permutations on a set Ω of v letters, and take an integer k with $1 < k < v$. Then G induces a group of permutations on the $\binom{v}{k}$ k -element subsets of Ω . The elements of Ω are the points and the k -element subsets of Ω the blocks of a (trivial) BIBD with parameters

$$v, b = \binom{v}{k}, \quad r = \binom{v-1}{k-1}, \quad k, \lambda = \binom{v-2}{k-2}$$

Here $r - \lambda = \binom{v-2}{k-1}$, and the following is a consequence of theorem 1.

Corollary 2. If G is a group of permutations on a set Ω of v letters and if p is a prime not dividing $\binom{v-2}{k-1}$, then there

is a strong p -semi-symmetry of the set of orbits of G on Ω to the set of orbits of G on the k -element subsets of Ω , and also a p -symmetry if $p \nmid (v, k)$ and $v_p(k) \geq v_p\left(\binom{v-1}{k-1}\right)$, and strong p -symmetry if $p \nmid k \binom{v-1}{k-1}$.

In particular, for $k = 2$ the conditions on the prime p for a p -symmetry (resp. strong p -semi-symmetry) are that $p \nmid (v-2)(v-1)$ (resp. $p \nmid v-2$) when $p > 2$, and $v \equiv 3 \pmod{4}$ (resp. $v \equiv 1 \pmod{2}$) when $p = 2$; and for $k = 3$ that $p \nmid (v-3)(v-2)(v-1)$ (resp. $p \nmid (v-2)(v-3)$) when $p > 3$, $v \equiv 4$ or $7 \pmod{9}$ (resp. $v \equiv 1 \pmod{3}$) when $p = 3$, and $v \equiv 0 \pmod{4}$ (resp. $v \equiv 0$ or $1 \pmod{4}$) when $p = 2$. These conditions for p -symmetries when $k = 2, 3$ give strong p -symmetries if $p \neq k$.

Another application of theorem 1 is as follows. Let there be given a BIBD D , with parameters v, b, r, k, λ , a group G of automorphisms of D , and a configuration defined in terms of the incidence relation in D , involving k' points, such that all sets of k' points of D satisfying the configuration form the blocks of a BIBD D' with the same points as D . Then G induces a group of automorphisms of D' with the same action on points as G .

We now give some examples of allowable configurations and the corresponding parameters b', r', k', λ' of D' :

(1) s points on some block, u points not on the block (we count a set of $s + u$ points as a block of D' as many times as s of them occur together on a block of D not containing any of the remaining u points); here

$$b' = b \binom{k}{s} \binom{v-k}{u}, \quad r' = r \binom{k-1}{s-1} \binom{v-k}{u} + (b-r) \binom{k}{s} \binom{v-k-1}{u-1}, \quad k' = s + u$$

$$\lambda' = \lambda \binom{k-2}{s-2} \binom{v-k}{u} + 2(r-\lambda) \binom{k-1}{s-1} \binom{v-k}{u-1} + (b-2r+\lambda) \binom{k}{s} \binom{v-k}{u-2}$$

where

$$\binom{a}{0} = 1, \quad \binom{a}{-1} = \binom{a}{-2} = 0$$

(2) *Quadrangles* (i.e., four points, no three on a block), where $\lambda = 1$; here

$$b' = \frac{v(v-1)(v-k)(v-3k+3)}{4!}, \quad r' = \frac{4b'}{v}, \quad k' = 4, \quad \lambda' = \frac{3r'}{v-1}$$

For protective planes one can similarly obtain the designs of all pentagons and of all hexagons.

When the configuration consists of $k - 1$ points on a block, we can identify the blocks of D' with incident point-block pairs in D , i.e., with *flags*. For a projective (resp. affine) plane of order n ,

$$\begin{aligned} r' - \lambda' &= n^2 + 1, & b' &= (n^2 + n + 1)(n + 1), & r' &= (n + 1)n, & k' &= n \\ (\text{resp. } r' - \lambda' &= n^2 - n + 1, & b' &= n^2(n + 1), & r' &= (n + 1)(n - 1), & k' &= n - 1) \end{aligned}$$

which gives the following result.

Corollary 3. Given a prime p and a group of automorphisms of a projective plane of order n , there is a strong p -semi-symmetry (resp. p -symmetry, strong p -symmetry) of the point (or line) orbits into the flag orbits provided:

$$p \nmid n^2 + 1 \text{ (resp. } p \nmid (n^2 + 1)(n + 1), p \nmid (n^2 + 1)(n + 1)n) \quad (6)$$

For an affine plane the conditions are

$$p \nmid n^2 - n + 1 \text{ (resp. } p \nmid (n^2 - n + 1)(n + 1)n, p \nmid (n^2 - n + 1)(n + 1)(n - 1)n) \quad (7)$$

When the configuration consists of k points on a block and one point not on the block, we can identify the blocks of D' with non-incident point-block pairs in D , i.e., with *anti-flags*. Here for a projective (resp. affine) plane of order n ,

$$\begin{aligned} r' - \lambda' &= n^2(n + 1), & b' &= n^2(n^2 + n + 1), & r' &= n^2(n + 2), & k' &= n + 2 \\ (\text{resp. } r' - \lambda' &= (n - 1)(n^2 + n + 1), & b' &= n^2(n + 1)(n - 1), & r' &= (n + 1)^2(n - 1), & k' &= n + 1) \end{aligned}$$

which gives the following result:

Corollary 4. The statement of corollary 2 holds for anti-flags in place of flags if Eq. (6) is replaced by

$$p \nmid n(n + 1) \text{ (resp. } p \nmid n(n + 1) \text{ and } p \neq 3, p \nmid n(n + 1)(n + 2))$$

and Eq. (7) is replaced by

$$p \nmid (n - 1)(n^2 + n + 1) \text{ (resp. } p \nmid (n - 1)(n^2 + n + 1)(n + 1), p \nmid (n - 1)(n^2 + n + 1)(n + 1))$$

Thus, these results give information even when p divides n .

References

1. Hall, M., Jr., *Combinatorial Theory*. Blaisdel Publishing Co., Waltham, Mass., 1967.
2. Block, R. E., "On the Orbits of Collineation Groups," *Math. Zeit.*, Vol. 96, pp. 33-49, 1967.
3. Dembowski, P., "Verallgemeinerungen von Transitivitätsklassen endlicher projektiver Ebenen," *Math. Zeit.*, Vol. 69, pp. 59-89, 1958.
4. Hughes, D. R., "Collineations and Generalized Incidence Matrices," *Trans. Am. Math. Soc.*, Vol. 86, pp. 284-296, 1957.
5. Parker, E. T., "On Collineations of Symmetric Designs," *Proc. Am. Math. Soc.*, Vol. 8, pp. 350-351, 1957.
6. De Bruijn, N. G., "Polya's Theorem of Counting," in *Applied Combinatorial Mathematics*, chap. 5. Edited by E. F. Beckenbach. John Wiley & Sons, New York, 1964.

E. Phase-Locking To Noisy Oscillators,

R. C. Tausworthe

1. Introduction

In most analyses of phase-lock operation, it is assumed that the loop is required to track a (modulated or unmodulated) *spectrally pure* source, and such analyses have resulted in worthwhile, accurate predictions of loop performance, so long as the spectral impurities were well within the loop bandwidth. Other analyses have considered the effects of noises in the oscillators themselves, and several spectral models have evolved, or been conjectured, and these, for the most part, point up several

inconsistencies between theoretical calculations and measurement techniques.

The most popular, accurate, and perhaps most easily implementable frequency measurement technique is indicated in Fig. 5: Two oscillators of comparable quality are heterodyned together, at a slight frequency offset, so that differential variations in the beat-frequency can be measured by cycle-counting. As one would expect, the measured variations decrease in magnitude as the measurement period increases, the specific function so obtained being related to the spectral density of the perturbing process.

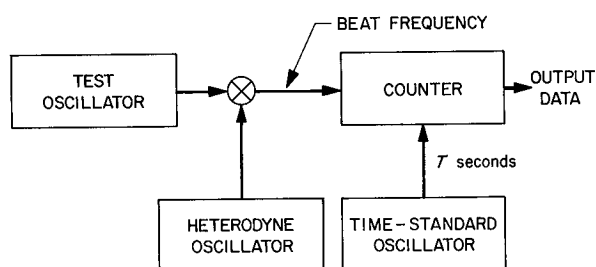


Fig. 5. Oscillator frequency stability measurement technique

Usually, the largest contributor by far to oscillator frequency-noise is the so-called "flicker" component, which, insofar as anyone has yet been able to determine experimentally, exhibits a $1/f$ characteristic all the way from the upper frequencies at which it becomes masked in other oscillator noises, down to the lowest measurable frequencies—less than 10^{-7} Hz.

But if one assumes this $1/f$ behavior is true at all frequencies, he is led to the theoretical conclusion that eventual frequency deviations have infinite variance, even though the average variation of measured finite-length samples is always bounded. A further consequence of the assumption is that slow deviations become larger in magnitude as their drift-rate decreases—a condition that contradicts the observed fact that the drift interval is limited to some small fraction of the oscillator center frequency, depending on the Q of the circuit.

If these two oscillators are placed in a phase-locked loop, one as a source and one as a VCO as shown in Fig. 6, the same assumption leads to a theoretical infinite loop phase-error variance (on an ensemble average, the usual statistical technique) for all loops, regardless of bandwidth, except for those having a perfect integrator

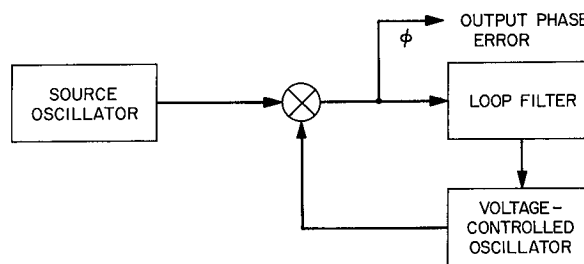


Fig. 6. Phase-locked loop frequency-stability measurement

in the loop filter. Then it is finite. The engineering fact, however, is that loops of the former type *do* lock and *do* track, with finite phase error whenever their bandwidth is sufficiently large.

Extrapolating the $1/f$ characteristic down to $f = 0$, it seems, is just too drastic an assumption, and, insofar as observable results are concerned, one which produces unrealistic theoretical results. Extrapolation of the $1/f$ behavior to infinite frequency does not yield any appreciable anomaly, however.

Accepting the frequency-counting technique as a calibration method for oscillators, one then needs a method of predicting the performance of a phase-locked system using these oscillators. In order to correlate the two behaviors, one needs a spectral model of the oscillators which indeed explains all observable results. In what follows, then, we shall investigate the effects of oscillator noises both on counted-frequency measurements and on loop error variances, and we shall develop for second-order loops, at least, a way to use counter data directly to produce predictions of loop error and drift rates.

2. The Noise Model

Rather than assuming that the $1/f$ -law holds at all frequencies, we shall assume that it holds only down to an angular frequency $\omega = \epsilon$ and that, at lower frequencies, the spectral density levels off to $1/\epsilon$. Such an assumption certainly fits in the observable region, and seems to restore the hope of a model producing potentially workable answers, but it conceivably introduces the need for knowing ϵ , a quantity not measurable by the frequency-counter technique, and not directly observable any other way.

We shall model the frequency disturbance of both oscillators by an equivalent noise $n(t)$ referred to the input of a perfect, unity-gain VCO, as in Fig. 7. The

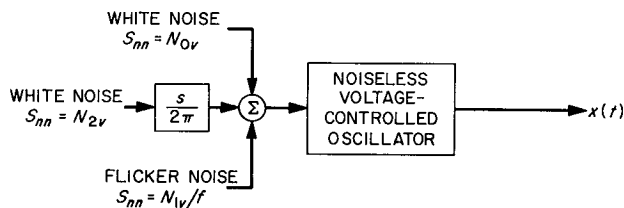
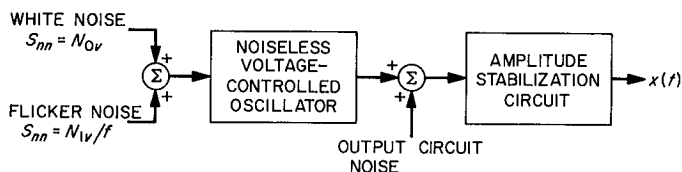


Fig. 7. The oscillator noise model

second oscillator will be assumed noiseless. The beat-note thus takes the form

$$x(t) = A(2)^{1/2} \cos [\omega_0 t + \theta_0 + \int^t n(t) dt] \quad (1)$$

$$S_{nn}(j2\pi f) = \begin{cases} \left[N_{0v} + \frac{2\pi N_{1v}}{\epsilon} + f^2 N_{2v} \right] |G(j2\pi f)|^2, & |f| < \epsilon/2\pi \\ \left[N_{0v} + \frac{N_{1v}}{|f|} + f^2 N_{2v} \right] |G(j2\pi f)|^2, & |f| > \epsilon/2\pi \end{cases} \quad (3)$$

Due either to the oscillator output circuit, or the loop- or counter-input circuit, there is a high-frequency cut-off effect which we have represented by the relatively wideband filter $G(s)$.

Concerning the bandwidths of various filters, such as $G(s)$, we shall define for the filter, say $U(s)$, the parameter w_U by the integral

$$w_U = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |U(j\omega)|^2 d\omega \quad (4)$$

when $U(0)$ is normalized to unity, w_U is the *fiducial bandwidth* (Ref. 1, Chap. 2) of U , and when $\max_{\omega} |U(j\omega)|$ is normalized to unity, w_U is the ordinary *equivalent noise bandwidth* of $U(s)$.

3. Counted-Frequency Measurements

Let us suppose that the counter configuration has produced n samples of data, each of duration T seconds.

in which A is the rms amplitude, ω_0 the mean beat-frequency, and θ_0 a uniformly distributed random phase.

There are three important noise components which comprise the output phase process: (1) There is a white-noise component introduced at the VCO input; (2) a "1/f" component at the VCO input; and (3) a wideband component at the oscillator output. The first of these is typical of thermal noises generated in resistors in the oscillator circuit; the second from noise commonly found in transistors, varactor diodes, and carbon resistors; and the third is again thermal noise $v(t)$, but at the oscillator output, appearing as

$$\begin{aligned} x(t) &= A_1(2)^{1/2} \cos [\omega_0 t + \theta_1(t)] + v(t) \\ &= A(t)(2)^{1/2} \cos [\omega_0 t + \theta(t)] \end{aligned} \quad (2)$$

the slight amplitude deviations are often compensated for, so $A(t)$ is effectively a constant A . The contribution of $v(t)$ to $\theta(t)$ is then approximately $v(t)/A$. As a result, the spectral density of the disturbances, referred to the source-VCO input, is of the form

Except for a negligible fractional-cycle roundoff error, the counter will register a sample frequency of

$$\omega_k = \omega_0 + \frac{\theta(kT) - \theta[(k-1)T]}{T} \quad (5)$$

at the end of the interval $[(k-1)T, kT]$.

It must be remembered that ω_0 is not directly observable; hence it is necessary to estimate it by

$$\hat{\omega}_0 = \frac{1}{n} \sum_{k=1}^n \omega_k \quad (6)$$

and the variance of this estimate by

$$\hat{\sigma}_\omega^2 = \frac{1}{n} \sum_{k=1}^n (\omega_k - \hat{\omega}_0)^2 \quad (7)$$

The actual (i.e., ensemble average) values are

$$\begin{aligned}\omega_0 &= E[(\hat{\omega}_0)] = E(\omega_k) \\ \sigma_\omega^2 &= E[(\omega_k - \omega_0)^2]\end{aligned}\quad (8)$$

While it is true that $\hat{\omega}_0$ is an unbiased estimator of ω_0 [i.e., $E(\hat{\omega}_0) = \omega_0$], it is not true that $\hat{\sigma}_\omega^2$ is unbiased. It is shown in the next article in this Section,³ that whenever $\epsilon T \ll 1$

$$\sigma_\omega^2 = \frac{N_{2v} w_G \Delta_G(T)}{2\pi^2 T^2} + \frac{N_{0v}}{T} + N_{1v} \left[5 - 2\gamma + 2 \ln \left(\frac{1}{\epsilon T} \right) \right] \quad (9)$$

where γ is Euler's constant, $\gamma = 0.5772 \dots$, and $\Delta_G(T) = 1 - R_{gg}(T)/w_G$. The mean sample-variance, under the same assumptions, with $n\epsilon T \ll 1$ in addition, is

$$\begin{aligned}E(\hat{\sigma}_\omega^2) &= \frac{N_{2v} w_G}{2\pi^2 T^2} \left[\frac{\Delta_G(T) - \Delta_G(nT)}{n^2} \right] \\ &\quad + \frac{(n-1)N_{0v}}{nT} + 2N_{1v} \ln(n)\end{aligned}\quad (10)$$

It is interesting to note for these moderately large n that $E(\hat{\sigma}_\omega^2)$ is independent of ϵ and that the flutter-noise component is independent of T , whereas the actual σ_ω^2 depends on both.

In the limit as $n \rightarrow \infty$ with $\epsilon T \ll 1$, it is further shown that $E(\hat{\sigma}_\omega^2)$ increases with n , asymptotically reaching σ_ω^2 , as the law of large numbers requires. Experimental observations with nT as large as 10^7 fail to produce a significant departure from the $\ln(n)$ approximation. It may be inferred from this then that ϵ must be smaller than 10^{-7} .

The common counter-instrumentation techniques usually result in a plot roughly similar to that in Fig. 8. There are three distinct regions, one corresponding to each of the three types of noise. At low values of T , the oscillator output-circuit noise (or counter input-circuit) noise seems to be the dominant contributor to $\hat{\sigma}_\omega^2$, and, depending on the cut-off characteristic $\Delta_G(T)$, the behavior commonly appears as anything from $1/T$ to $1/T^2$.

In the next region, the oscillator internal white noise takes predominance, and $\hat{\sigma}_\omega^2$ varies as $1/T$. Finally, as T grows larger, the flicker noise becomes evident, and $\hat{\sigma}_\omega^2$ depends significantly upon the number n of sample points,

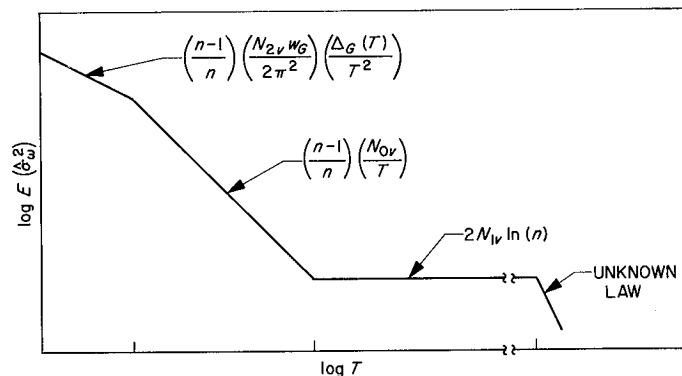


Fig. 8. The form of counted-frequency oscillator-stability data

but not upon T or ϵ . Finally, as T becomes larger than $1/\epsilon$, the behavior must again decrease.

The difference between σ_ω^2 and $E(\hat{\sigma}_\omega^2)$ is precisely equal to $E[(\hat{\omega}_0 - \omega_0)^2]$, and thus it is evident, due to the extreme smallness of ϵ , that $E[(\hat{\omega}_0 - \omega_0)^2]$ may consequently be relatively large. What this means in terms of oscillator instability is that the estimated center frequency has a much larger variance than do the shorter-term fluctuations about this estimation. It is common to refer to $E[(\hat{\omega}_0 - \omega_0)^2]$ as the *long-term drift*, and $E(\hat{\sigma}_\omega^2)$ as the *short-term frequency variation*.

4. Loop Tracking Errors

Now let us suppose that a phase-locked loop has acquired and is tracking the oscillator in Part 2 above (Fig. 6). The phase error, in the absence of loop input noise is given by

$$\phi(t) = \left[\frac{1 - L(p)}{p} \right] n(t) \quad (11)$$

where $L(s)$ is the ordinary loop transfer function (Ref. 1, Chap. 5). It is convenient to define

$$H(s) = \frac{1 - L(s)}{s} \quad (12)$$

as the frequency-error transfer characteristic. The error at any time t is then the convolution

$$\phi(t) = \int_0^\infty h(u) n(t-u) du \quad (13)$$

in which $h(\tau)$ is the unit-impulse response of $H(s)$. It thus follows directly that the total mean-square phase

³F. Analysis of the Effect of Input Noise on a VCO, R. Gray and R. C. Tausworthe.

error is

$$\begin{aligned}\sigma_{\phi}^2 &= \frac{1}{\pi} \int_0^{+\infty} |H(j\omega)|^2 S_{nn}(j\omega) d\omega \\ &= \frac{1}{\pi} \int_0^{\epsilon} |H(j\omega)|^2 S_{nn}(j\omega) d\omega + \frac{1}{\pi} \int_{\epsilon}^{\infty} |H(j\omega)|^2 S_{nn}(j\omega) d\omega\end{aligned}\quad (14)$$

One normally apportions the total error into a steady-state component caused by having mismatched VCO center frequencies, and a jitter component superimposed. After a given time T , the usual estimate of this steady-state phase error is

$$\hat{\phi}_{ss} = \frac{1}{T} \int_0^T \phi(t) dt \quad (15)$$

and the variance-estimate $\hat{\sigma}_{\phi}^2$ of the loop error about this $\hat{\phi}_{ss}$ is then an expression whose mean value is of the form

$$E(\hat{\sigma}_{\phi}^2) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |H(j\omega)|^2 S_{nn}(j\omega) \left\{ 1 - \left[\frac{\sin \frac{\omega T}{2}}{\frac{\omega T}{2}} \right]^2 \right\} d\omega \quad (16)$$

The value $E(\hat{\sigma}_{\phi}^2)$ is the same as σ_{ϕ}^2 for a loop tuned every T seconds to remove the apparent steady-state phase offset. The term in braces has a double zero at the origin, so that $E(\hat{\sigma}_{\phi}^2)$ would converge even if $S_{nn}(j\omega)$ were to have a true $1/f$ term. In fact, for $0 \ll T \ll 1/\epsilon$ the term in braces can be approximated very closely by a unit-step function at $\omega = 2\pi/T$, so that

$$E(\hat{\sigma}_{\phi}^2) \approx \frac{1}{\pi} \int_{2\pi/T}^{\infty} |H(j\omega)|^2 S_{nn}(j\omega) d\omega \quad (17)$$

the mean estimated loop variance $\hat{\sigma}_{\phi}^2$ thus resembles the second term in Eq. (14) for the total loop error $\hat{\sigma}_{\phi}^2$, but with ϵ replaced by $2\pi/T$. For very large T , phase errors arising from the wideband noises are the same, whether the loop is tunable or not:

$$\sigma_{\phi}^2 = \sigma_{\phi T}^2 = N_{ov} w_{GH} + \frac{N_{2v}}{2\pi^2} w_{G(1-L)} \quad (\text{no flutter noise}) \quad (18)$$

The numbers w_{GH} and $w_{G(1-L)}$ are the bandwidth integrals

$$w_{UV} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |U(j\omega)|^2 |V(j\omega)|^2 d\omega \quad (19)$$

a. First-order loop. The transfer function of a first-order loop is (Ref. 1, Chap. 5)

$$L(s) = \frac{2w_L}{s + 2w_L} \quad (20)$$

in terms of the loop two-sided bandwidth w_L . The frequency-error characteristic is then

$$H(s) = \frac{1}{s + 2w_L} = \frac{1}{2w_L} L(s) \quad (21)$$

and if $w_G \gg w_L$ (the usual case), we have

$$\begin{aligned}w_{GH} &\approx \frac{1}{4w_L} \\ w_{G(1-L)} &\approx w_G\end{aligned}\quad (22)$$

The flicker component alone produces

$$\sigma_{\phi}^2 \approx \frac{N_{1v}}{2w_L^2} \left[1 + \ln \left(\frac{2w_L}{\epsilon} \right) \right] \quad (\text{flicker only})$$

$$E(\hat{\sigma}_{\phi}^2) \approx \frac{N_{1v}}{2w_L^2} \ln(w_L T) \quad (23)$$

Hence, the phase errors are of the form

$$\hat{\sigma}_{\phi}^2 = \frac{N_{2v} w_G}{2\pi^2} + \frac{N_{ov}}{4w_L} + \frac{N_{1v}}{2w_L^2} \left[1 + \ln \left(\frac{2w_L}{\epsilon} \right) \right] \quad (24)$$

and

$$E(\hat{\sigma}_{\phi}^2) = \frac{N_{2v} w_G}{2\pi^2} + \frac{N_{ov}}{4w_L} + \frac{N_{1v}}{2w_L^2} \ln(w_L T) \quad (25)$$

Note the dependence of σ_{ϕ}^2 in Eq. (24) on the unknown parameter ϵ . Just as σ_{ω}^2 is not measurable physically, neither is σ_{ϕ}^2 ; rather, there is a long-term steady-state phase drift, $E(\hat{\phi}_{ss}^2)$, and short-term fluctuations about it, $E(\hat{\sigma}_{\phi}^2)$.

b. Second-order loop. The second-order loop with filter

$$F(s) = \frac{1 + \tau_2 s}{1 + \tau_1 s} \quad (26)$$

has, as its frequency-error characteristic, the function

$$H(s) = \frac{s + \frac{1}{\tau_1}}{s^2 + 2\zeta\beta s + \beta^2}$$

in which ζ and β are the damping coefficient and natural frequency of the loop, respectively, as in Ref. 1, Chap. 5. Under the usual assumption $\tau_2 \ll r\tau_1$, these are approximately

$$\begin{aligned}\zeta &= \frac{1}{2} r^{1/2} \\ \beta &= \frac{r}{\tau_2}\end{aligned}\quad (27)$$

in terms of the loop parameter $r = AK\tau_2^2/\tau_1$, A being the rms input signal level and K , the open-loop gain. The total mean-square error integral now evaluates to approximately

$$\begin{aligned}\sigma_\phi^2 &= \frac{N_{2v}w_G}{2\pi^2} + \left(\frac{r+1}{4r}\right) \frac{N_{0v}}{w_L} + \left(\frac{1}{2}\left(\frac{r+1}{r}\right)^2 \left(\frac{\tau_2}{\tau_1}\right)^2\right. \\ &\quad \times \left. \left\{1 + \ln \left[\left(\frac{r}{r+1}\right) \frac{2w_L}{\epsilon}\right]\right\} + g(r)\right) \frac{N_{1v}}{w_L^2}\end{aligned}\quad (28)$$

whenever $\tau_2 \ll \tau_1$. The function $g(r)$ is given in Eq. (5-28) of Ref. 1. As $\tau_2/\tau_1 \rightarrow 0$ at a fixed value of r , Eq. (28) agrees as it should with that given in Eq. (5-27) of Ref. 1 for the perfect-integrator loop.

Here it is interesting to note that the term containing ϵ is negligible whenever ϵ satisfies the condition

$$\epsilon \gg 2w_L \left(\frac{r}{r+1}\right) \exp \left[-2 \left(\frac{\tau_1}{\tau_2}\right)^2 \left(\frac{r}{r+1}\right)^2 g(r) \right]\quad (29)$$

At critical damping ($r = 4$ and $g(4) = 1.5625$), the condition above states that the contribution of ϵ to σ_ϕ^2 will be negligible whenever

$$\epsilon \gg 1.8 w_L \exp \left[-2 \left(\frac{\tau_1}{\tau_2}\right)^2 \right]\quad (30)$$

A modest value of τ_2/τ_1 of 100 indicates that, in order to be significant, ϵ must be less than 10^{-8680} ; and while it is not known how small a typical ϵ actually is, one feels intuitively that it surely cannot be anywhere near as small as 10^{-8680} .

The total phase error is therefore

$$\sigma_\phi^2 = \frac{N_{2v}w_G}{2\pi^2} + \left(\frac{r+1}{4r}\right) \frac{N_{0v}}{w_L} + \frac{g(r)N_{1v}}{w_L^2}\quad (31)$$

This analysis shows that there is no essential difference between $\hat{\sigma}_\phi^2$ and $E(\sigma_\phi^2)$, as there is in the first-order loop.

5. Correlation Between Frequency-Count Measurements and Loop Phase Errors

Given the $\hat{\sigma}_\phi^2$ versus T plot of a counted-frequency dispersion graph for a specific known sample size n , one may find best-fit parameters N_{0v} and N_{1v} . If one further knows the characteristic $G(s)$, either from a familiarity with the oscillator-output or the counter-input circuits, he can also compute the characteristic $\Delta_G(t)$, and then, from it, the value N_{2v} . It is also possible to measure N_{2v} directly in many cases.

It is worthwhile to point out that since the characteristic $G(s)$ includes the combined effects of the oscillator output filter, the receiver input filter, and usually the receiver IF filter, the $G(s)$ used in loop-performance calculations is not the same as the one used in the counted-frequency model.

Once N_{0v} , N_{1v} , and N_{2v} are known, they can be substituted into the formulas for σ_ϕ^2 to produce performance predictions.

Reference

1. Tausworthe, R. C., *Theory and Practical Design of Phase-Locked Receivers*, Technical Report 32-819. Jet Propulsion Laboratory, Pasadena, Calif., February 15, 1966.

F. Analysis of the Effect of Input Noise on a VCO,

R. M. Gray and R. C. Tausworthe

1. Introduction

The variance and the expected value of the sample variance of the counted-frequency output of a perfect VCO are theoretically derived as a function of an equivalent input noise power spectral density $S_{nn}(j\omega)$. The results are applied to white noise and to "flicker" noise equivalent inputs. It is seen that the expected value of the sample variance is a useful tool for estimating the spectral parameters involved.

2. Derivation of Actual and Sample Variances

The block diagram of the system and the pertinent definitions are given in Fig. 9. The random variable of interest is the sample frequency

$$v_i = \frac{1}{T} \int_{(i-1)T}^{iT} n(t) dt = \frac{1}{T} [\phi(iT) - \phi((i-1)T)] \quad (1)$$

where $\phi(t)$ is the output phase of the VCO at time t .

We assume that the noise has mean zero and spectral density spectrum $S_{nn}(j\omega)$. The variance of the v_i is then

$$\begin{aligned} \sigma_{v_i}^2 &= E\{[v_i - E(v_i)]^2\} \\ &= E(v_i^2) - E^2(v_i) \\ &= E(v_i^2) \end{aligned} \quad (2)$$

since $E(v_i) = 0$. Combining Eqs. (1) and (2), we find that

$$\begin{aligned} \sigma_{v_i}^2 &= E\left[\left(\frac{1}{T} \int_{(i-1)T}^{iT} n(t) dt\right) \left(\frac{1}{T} \int_{(i-1)T}^{iT} n(t) dt\right)\right] \\ &= \frac{1}{T^2} \int_0^T d\alpha \int_0^T d\beta R_{nn}[\alpha + (i-1)T, \beta + (i-1)T] \end{aligned} \quad (3)$$

where $R_{nn}(t_1, t_2)$ is the ensemble autocorrelation of the noise $n(t)$. Presuming that the noise is at least wide-sense stationary, we then obtain

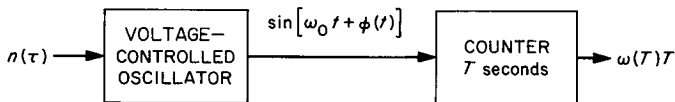
$$\begin{aligned} \sigma_{v_i}^2 &= \frac{1}{T^2} \int_0^T d\alpha \int_0^T d\beta R_{nn}(\alpha - \beta) \\ &= \frac{1}{T} \int_{-T}^T d\tau \left(1 - \frac{|\tau|}{T}\right) R_{nn}(\tau) \end{aligned} \quad (4)$$

a result which no longer depends on the interval index i . Next, we substitute

$$R_{nn}(\tau) = \int_{-\infty}^{\infty} S_{nn}(j\omega) e^{j\omega\tau} \frac{d\omega}{2\pi} \quad (5)$$

to arrive at the equation

$$\sigma_{v_i}^2 = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} S_{nn}(j\omega) \left(\frac{\sin \frac{\omega T}{2}}{\frac{\omega T}{2}}\right)^2 \quad (6)$$



$$\omega(T)T = \omega_0 T + \phi(T) - \phi(0)$$

$$\omega(T) = \frac{1}{T} [\phi(T) - \phi(0)] + \omega_0$$

$$v = \omega(T) - \omega(0) = \frac{1}{T} [\phi(T) - \phi(0)] = \frac{1}{T} \int_0^T n(t) dt$$

Fig. 9. System block diagram and definitions

This is the true variance of the VCO output counted frequency.

In a like manner, the sample variance, defined as

$$\hat{\sigma}_N^2 = \frac{1}{N} \sum_{i=1}^N v_i^2 - \left(\frac{1}{N} \sum_{i=1}^N v_i\right)^2 \quad (7)$$

is a random variable by which σ_v^2 can be estimated, since σ_v^2 is not observable directly. Whenever the law of large numbers applies, $\hat{\sigma}_N^2 \rightarrow \sigma_v^2$, if the latter exists. It is noteworthy that $\hat{\sigma}_N^2$ may exist, even if σ_v^2 does not, and represents the variation, or performance figure of the oscillator over a short term.

Since $\hat{\sigma}_N^2$ is itself a random variable, we can form its ensemble average

$$E(\hat{\sigma}_N^2) = \frac{1}{N} \sum_{i=1}^N E(v_i^2) - \frac{1}{N^2} \sum_{i=1}^N \sum_{l=1}^N E(v_i v_l) \quad (8)$$

Whenever the random variables v_i are uncorrelated Eq. (8) yields the well-known result

$$E(\hat{\sigma}_N^2) = \left(\frac{N-1}{N}\right) \sigma_v^2 \quad (9)$$

Note that the factor $N/(N-1)$ frequently used to make $\hat{\sigma}_N^2$ an unbiased estimator when the v_i are uncorrelated is not used here. Evaluating $E(v_i v_l)$, we find for the

more general, correlated case that

$$E(v_i v_l) = \frac{1}{T^2} \int_0^T d\alpha \int_0^T d\beta R_{nn}[\alpha + (i-1)T, \beta + (\ell-1)T] \quad (10)$$

Again, by assuming at least wide-sense stationarity, letting $\tau = \alpha - \beta$ and interchanging the order of the integration, we obtain the formula

$$E(\hat{\sigma}_N^2) = \frac{1}{T} \int_{-T}^T d\tau \left(1 - \frac{|\tau|}{T}\right) R_{nn}(\tau) - \frac{1}{N^2} \sum_{i=1}^N \sum_{l=1}^N \frac{1}{T} \int_{-T}^T d\tau \left(1 - \frac{|\tau|}{T}\right) R_{nn}[\tau + (i-l)T] \quad (11)$$

which then yields the mean sample-variance in terms of the perturbing spectrum:

$$E(\hat{\sigma}_N^2) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} S_{nn}(j\omega) \left(\frac{\sin \frac{\omega T}{2}}{\frac{\omega T}{2}} \right)^2 - \frac{1}{N^2} \sum_{i=1}^N \sum_{l=1}^N \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} S_{nn}(j\omega) e^{j\omega(i-l)T} \left(\frac{\sin \frac{\omega T}{2}}{\frac{\omega T}{2}} \right)^2 \quad (12)$$

Then, recognizing that

$$\sum_{i=1}^N \sum_{l=1}^N e^{j\omega(i-l)T} = \left(\frac{\sin \frac{N\omega T}{2}}{\sin \frac{\omega T}{2}} \right)^2 \quad (13)$$

we arrive at the final result

$$E(\hat{\sigma}_N^2) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} S_{nn}(j\omega) \left[\left(\frac{\sin \frac{\omega T}{2}}{\frac{\omega T}{2}} \right)^2 - \left(\frac{\sin \frac{N\omega T}{2}}{\frac{N\omega T}{2}} \right)^2 \right] \quad (14)$$

With $x = \omega T/2$, Eq. (14) takes the alternate form

$$E(\hat{\sigma}_N^2) = \frac{1}{\pi T} \int_{-\infty}^{\infty} dx S_{nn}\left(\frac{j2x}{T}\right) \left[\left(\frac{\sin x}{x} \right)^2 - \left(\frac{\sin nx}{nx} \right)^2 \right] \quad (15)$$

3. Evaluation for Flicker Noise

A case of special practical interest is that of so-called "flicker" noise having a spectrum behaving as $K/|\omega|$, where K is some constant. Any investigation of such a spectrum is complicated by the fact that $K/|\omega|$ is non-integrable and therefore cannot accurately model any real, finite-power process. Any physical process must necessarily have both low-frequency and high-frequency "cut-off" points, say ϵ and η , beyond which the spectrum ceases to behave as $K/|\omega|$. Even though such "cut-off" points must exist, experimental techniques have not yet been able to find them. Thus any such points must be treated as unknowns, along with K .

The two parameters of interest, σ_v^2 and $E(\hat{\sigma}_N^2)$, will be evaluated by two separate methods. As a first attempt

we will assume that $S_{nn}(j\omega) = K/|\omega|$ for all ω , and simply substitute this relation into Eqs. (6) and (15). Such an approach is mathematically simple, but it leaves some questions unanswered since the model itself is physically unrealistic. As a second more realistic and more rigorous approach, we shall take the "cut-off" points into account by using the $S_{nn}(j\omega)$ pictured in Fig. 10. We shall then allow the high-frequency cut-off to approach infinity, but shall require $\epsilon > 0$. It will be seen that the two approaches yield consistent results.

a. First method. Merely plugging $S_n(j\omega) = K/|\omega|$ into Eqs. (6) and (15) requires some justification. Since $K/|\omega|$ has no Fourier transform, it can be argued that the arguments leading to those equations which involve $R_{nn}(\tau)$ are invalid. Since the justification of the derivation

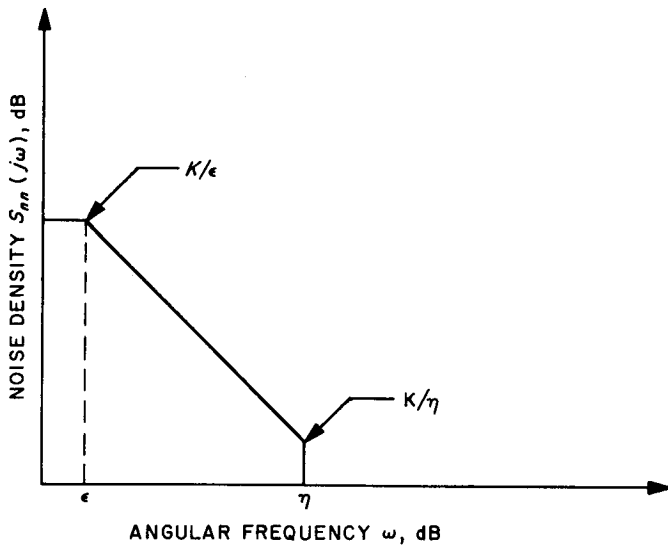


Fig. 10. Spectral density model of oscillator noise

involves arguments essentially the same as those used in the second method, with the exception that ϵ is allowed to go to zero, we will here simply assume that Eqs. (6) and (15) are valid when $S_{nn}(j\omega) = K/|\omega|$ as a limiting case of the second method, and postpone the needed rigor to a later discussion.

Substituting $S_{nn}(j\omega) = K/|\omega|$ into Eqs. (6) and (15), we find

$$\sigma_v^2 = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{K}{|\omega|} \left(\frac{\sin \frac{\omega T}{2}}{\frac{\omega T}{2}} \right)^2 \quad (16)$$

$$E(\hat{\sigma}_N^2) = \int_{-\infty}^{\infty} dx \frac{1}{x} \left[\left(\frac{\sin x}{x} \right)^2 - \left(\frac{\sin nx}{nx} \right)^2 \right] \quad (17)$$

Equation (16) demonstrates the type of problem involved in the analysis of flicker noise: The σ_v^2 integral does not exist; but even though σ_v^2 is infinite, $E(\hat{\sigma}_N^2)$ is finite, independent of T , and, as shown in Table 3, takes the value

$$E(\hat{\sigma}_N^2) = \frac{K}{\pi} \ln N \quad (18)$$

Since it is the sample variance and not the actual variance which one measures in an experiment, the infinite actual variance does not prevent us from determining the value of K , given some estimate of $E(\hat{\sigma}_N^2)$ for a specific N .

Table 3. Evaluation of integral

$$f(n) = \int_a^b \frac{dx}{x} [g(x) - g(nx)]$$

By differentiating with respect to n ,

$$f'(n) = - \int_a^b dx g'(nx)$$

Letting $nx = t$ yields

$$\begin{aligned} f'(n) &= - \frac{1}{n} \int_{na}^{nb} g'(t) dt \\ &= \frac{g(na) - g(nb)}{n} \end{aligned}$$

If

$$g(x) = \frac{K}{\pi} \left(\frac{\sin x}{x} \right)^2, \quad a = 0, \quad b = \infty$$

then integrating with respect to n yields

$$f(n) = I + \frac{K}{\pi} \ln n [g(0) - g(\infty)] = \frac{K}{\pi} \ln n + I$$

where I is independent of N .

Table 4. Autocorrelation function of truncated $1/f$ spectrum

$$R_{nn}(\tau) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} S_{nn}(j\omega) e^{j\omega\tau}$$

$$= \int_0^{\infty} \frac{d\omega}{\pi} S_{nn}(j\omega) \cos \omega\tau$$

since $S_{nn}(j\omega)$ is an even function. Substituting for $S_{nn}(j\omega)$,

$$R_{nn}(\tau) = \int_0^{\epsilon} \frac{d\omega}{\pi} \frac{K}{\omega} \cos \omega\tau + \int_{\epsilon}^{\eta} \frac{d\omega}{\pi} \frac{K}{\omega} \cos \omega\tau$$

Assume $\epsilon \ll 1/\tau$ and $\eta \rightarrow \infty$, then

$$\begin{aligned} R_{nn}(\tau) &= \frac{K}{\pi} + \frac{K}{\pi} \int_{\epsilon}^{\infty} d\omega \frac{\cos \omega\tau}{\omega} \\ &= \frac{K}{\pi} \left(1 + \int_{\epsilon\tau}^{\infty} dx \frac{\cos x}{x} \right) \end{aligned}$$

At small values of $\epsilon\tau$ this reduces to

$$R_{nn}(\tau) = \frac{K}{\pi} \left(1 - \gamma + \ln \frac{1}{\epsilon\tau} \right)$$

where γ is Euler's constant.

The expected value of the sample variance is finite despite its derivation from an infinite variance process because the mean-frequency-estimation operation acts as a high-pass filter, canceling the infinite low-frequency power of the flicker noise. Since

$$\lim_{n \rightarrow \infty} E(\hat{\sigma}_N^2)$$

is infinite, one might say that the sample variance "converges" to the actual variance in some sense as N goes to infinity.

b. Second method. In order to make more rigorous the results of Part *a*, above, it is necessary to examine the model more closely, reflecting upon an actual process. The spectral model, as we have previously indicated, is shown in Fig. 10. Although actual flicker noise probably does not exhibit the sharp cut-offs assumed in the model, the evaluation of statistics using the model will give approximate results if η and ϵ are chosen properly.

The autocorrelation for this process exists and is shown in Table 4 to be

$$R_{nn}(\tau) = \frac{K}{\pi} \left[1 - \gamma + \ln \left(\frac{1}{\epsilon \tau} \right) \right] \quad (19)$$

It has been assumed that $\eta \rightarrow \infty$, $0 < \epsilon < 1/T$, $\tau < 1/\epsilon$, and γ is Euler's constant. The assumption that $\eta \rightarrow \infty$ will be used in all results of this method since even for moderately large η , the parameters of interest are both essentially independent of η . Since there is a well-defined autocorrelation, all the results of Subsection 2 apply to this spectrum.

The variance of this process is

$$\sigma_v^2 = 2 \left[\int_0^\epsilon \frac{d\omega}{2\pi} \frac{K}{\epsilon} \left(\frac{\sin \frac{\omega T}{2}}{\frac{\omega T}{2}} \right)^2 + \int_\epsilon^\eta \frac{d\omega}{2\pi} \frac{K}{\omega} \left(\frac{\sin \frac{\omega T}{2}}{\frac{\omega T}{2}} \right)^2 \right] \quad (20)$$

The assumption $\epsilon < 1/T$ provides

$$\int_0^\epsilon \frac{d\omega}{2\pi} \frac{K}{\epsilon} \left(\frac{\sin \frac{\omega T}{2}}{\frac{\omega T}{2}} \right)^2 \approx \int_0^\epsilon \frac{d\omega}{2\pi} \frac{K}{\epsilon} = \frac{K}{2\pi} \quad (21)$$

Evaluation of the second integral gives

$$\begin{aligned} \int_\epsilon^\eta \frac{d\omega}{2\pi} \frac{K}{\omega} \left(\frac{\sin \frac{\omega T}{2}}{\frac{\omega T}{2}} \right)^2 &= \frac{K}{2\pi} \left[\frac{1 - \cos T}{(\epsilon T)^2} - \frac{1 - \cos \eta T}{(\eta T)^2} \right. \\ &\quad \left. + \frac{\sin \epsilon T}{\epsilon T} - \frac{\sin \eta T}{\eta T} + \int_{\epsilon T}^{\eta T} \frac{\cos x}{x} dx \right] \quad (22) \end{aligned}$$

which, as η goes to infinity, produces the result

$$\sigma_v^2 = \frac{5}{2} \frac{K}{\pi} + \frac{K}{\pi} \int_{\epsilon T}^\infty \frac{\cos x}{x} dx \quad (23)$$

But since $\epsilon \ll 1$ by assumption, Eq. (23) is approximately

$$\sigma_v^2 = \frac{K}{\pi} \left(\frac{5}{2} - \gamma + \ln \frac{1}{\epsilon T} \right) \quad (24)$$

We see from this result that the variance is strongly dependent on ϵ and goes to infinity as ϵ becomes zero.

The variance decreases as T grows, however. Equation (24) is invalid if T becomes too large, and it can be seen from Eq. (6) that σ_v^2 goes to zero as T goes to infinity for any $\epsilon > 0$. This behavior is what one would expect, as frequency-counting followed by division by T is an averaging process and it is generally true that time averages of wide-sense stationary processes converge in the mean.

Substituting the present $S_{nn}(j\omega)$ into the sample-variance equation, we obtain

$$\begin{aligned} E(\hat{\sigma}_N^2) &= \frac{2}{\pi T} \int_0^{\epsilon T/2} dx \frac{K}{\epsilon} \left[\left(\frac{\sin x}{x} \right)^2 - \left(\frac{\sin Nx}{Nx} \right)^2 \right] \\ &\quad + \frac{K}{\pi} \int_{\epsilon T/2}^{\eta T/2} \frac{dx}{x} \left[\left(\frac{\sin x}{x} \right)^2 - \left(\frac{\sin Nx}{Nx} \right)^2 \right] \quad (25) \end{aligned}$$

Evaluation of the first term shows that it is $O(\epsilon T)^2$, and thus, for $(\epsilon T) \ll 1$, is negligible. The second term is evaluated by the method shown in Table 3 to give

$$\begin{aligned} E(\hat{\sigma}_N^2) &= \frac{K}{\pi} \left[\frac{1 - \cos \epsilon T}{(\epsilon T)^2} - \frac{1 - \cos N\epsilon T}{(N\epsilon T)^2} + \frac{\sin \epsilon T}{\epsilon T} \right. \\ &\quad \left. - \frac{\sin N\epsilon T}{N\epsilon T} + \int_{\epsilon T}^{N\epsilon T} \frac{\cos x}{x} dx \right] + I \quad (26) \end{aligned}$$

where I is a function independent of N . Under the additional assumption that N is only moderately large, so that $N\epsilon T \ll 1$, then Eq. (26) reduces to

$$E(\hat{\sigma}_N^2) = \frac{K}{\pi} \int_{\epsilon T}^{N\epsilon T} \frac{dx}{x} + I \approx \frac{K}{\pi} \ln N + I \quad (27)$$

But since

$$E(\hat{\sigma}_1^2) = E(V_1^2 - V_1^2) = 0$$

I must be zero; thus we have

$$E(\hat{\sigma}_N^2) = \frac{K}{\pi} \ln N \quad (28)$$

Thus if $0 < \epsilon < 1/T$, $N \ll 1/\epsilon T$, we have the same result as that given in Eq. (18). This says that for a small enough cut-off and a moderate N , $E(\hat{\sigma}_N^2)$ is independent of both T and ϵ and is the same (except for negligible terms in $N\epsilon T$ and ϵT) as the value of $E(\hat{\sigma}_N^2)$ for "true" $K/|\omega|$ noise. The only difference between Eqs. (18) and (28) is the "moderate N " restriction on the latter. Removing this restriction and re-evaluating Eq. (26) for $\epsilon T \ll 1$ gives

$$\begin{aligned} E(\hat{\sigma}_N^2) &= \frac{2}{\pi T} \int_0^{\epsilon T/2} dx \frac{K}{\epsilon} \left[\left(\frac{\sin x}{x} \right)^2 - \left(\frac{\sin Nx}{Nx} \right)^2 \right] \\ &+ \frac{K}{\pi} \left[\frac{1 - \cos \epsilon T}{(\epsilon T)^2} - \frac{1 - \cos N\epsilon T}{(N\epsilon T)^2} \right. \\ &+ \frac{\sin \epsilon T}{\epsilon T} - \frac{\sin N\epsilon T}{(N\epsilon T)} + \left. \int_{\epsilon T}^{N\epsilon T} \frac{\cos x}{x} dx \right] \\ &\rightarrow \frac{K}{\pi} \left(\frac{5}{2} + \int_{\epsilon T}^{\infty} \frac{\cos x}{x} dx \right) \end{aligned} \quad (29)$$

and hence in the limit

$$\lim_{N \rightarrow \infty} E(\hat{\sigma}_N^2) = \frac{K}{\pi} \left(\frac{5}{2} - \gamma + \ln \frac{1}{\epsilon T} \right) = \sigma_v^2 \quad (30)$$

That is, $E(\hat{\sigma}_N^2)$ approaches the actual variance σ_v^2 as N goes to infinity, even though it is a strong function of N .

The two approaches used to evaluate σ_v^2 and $E(\hat{\sigma}_N^2)$ are consistent and demonstrate the independence of $E(\hat{\sigma}_N^2)$ on ϵ for moderate N and the strong dependence of σ_v^2 on N and upon ϵ for very large N . One may thus estimate K from an experimental estimate of $E(\hat{\sigma}_N^2)$ at a moderate N . If one can find an NT large enough that $E(\hat{\sigma}_N^2)$ ceases to be a function of N , then it is possible that one may be able to estimate ϵ as well.

4. Evaluation for White Noise

Evaluating Eqs. (6) and (15) for the case of $S_n(j\omega) = N_0/2$ gives

$$\sigma_v^2 = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{N_0}{2} \left(\frac{\sin \frac{\omega T}{2}}{\frac{\omega T}{2}} \right)^2 = \frac{N_0}{2T} \quad (31)$$

$$\begin{aligned} E(\hat{\sigma}_N^2) &= \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{N_0}{2} \left[\left(\frac{\sin \frac{\omega T}{2}}{\frac{\omega T}{2}} \right)^2 - \left(\frac{\sin \frac{N\omega T}{2}}{\frac{N\omega T}{2}} \right)^2 \right] \\ &= \frac{N_0}{2T} \left(\frac{N-1}{N} \right) \end{aligned} \quad (32)$$

These results are well known and can be seen to be an example of Eq. (9).

5. Flicker Noise and White Noise

When flicker noise and white noise occur together, the white noise will predominate for small T and the flicker noise for large T . A sketch of $E(\hat{\sigma}_N^2)$ as a function of T is given in Fig. 11.

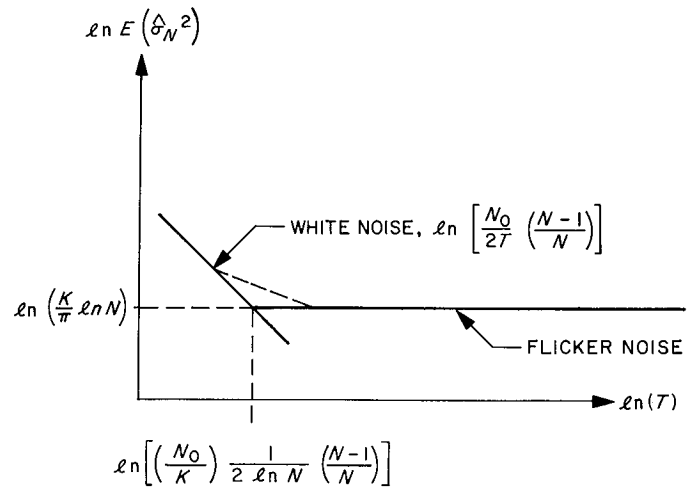


Fig. 11. Sketch of $E(\hat{\sigma}_N^2)$ in the presence of both white and flicker noise

6. Dead-Time in Experiment

A counter used in finding $\phi(T)$ may have a short period of time wherein it is through counting the last sample and not yet started on the next. Such may be due to the time required to print the count, for example. The preceding analysis changes slightly, because the sampling intervals are no longer exactly adjacent. If the dead-time is τ and the total sample time is T , we must redefine the random variable

$$v_i = \frac{1}{T - \tau} \int_{(i-1)T}^{iT - \tau} n(\tau) d\tau \quad (33)$$

Then by the previous procedure this leads to a revised $E(\hat{\sigma}_N^2)$:

$$E(\hat{\sigma}_N^2) = \frac{1}{\pi T} \int_{-\infty}^{\infty} dx S_n \left(j \frac{2x}{\rho T} \right) \left[\left(\frac{\sin x}{x} \right)^2 - \left(\frac{\sin \frac{N}{\rho} x}{\frac{N}{\rho} x} \right)^2 \right] \quad (34)$$

where $\rho = 1 - \tau/T$. For flicker noise this yields

$$E(\hat{\sigma}_N^2) = \frac{K}{\pi} \ln \frac{N}{\rho} \quad (35)$$

For white noise Eq. (32) gives

$$E(\hat{\sigma}_N^2) = \frac{N_0}{2T} \left(\frac{N - \rho}{N} \right)$$

and the resulting effects are seen to be slight when $\rho \approx 1$.

7. White Noise in the Oscillator Output Circuit

A small wideband noise in the oscillator output combines to produce an equivalent wideband noise in the oscillator output phase. Referred to the VCO input this appears as $d\nu/dt$, so the spectrum of this component is $S_{nn}(j\omega) = \omega^2 S_{vv}(j\omega)$. Thus, according to Eq. (6)

$$\begin{aligned} \sigma_v^2 &= \frac{2}{T^2} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} S_{vv}(j\omega) (1 - \cos \omega T) \\ &= \frac{2}{T^2} [R_{vv}(0) - R_{vv}(T)] \end{aligned} \quad (36)$$

where $R_{vv}(T)$ is the autocorrelation of the oscillator output wideband noise process. By defining $R_{vv}(0) = w_v J$, i.e., a noise bandwidth w_v times a noise density J , we have

$$\sigma_v^2 = 2Jw_v \frac{\Delta_v(T)}{T^2} \quad (37)$$

with $\Delta_v(T) = 1 - R_{vv}(T)/w_v J$. Evaluation of Eq. (15) for this case produces, for the mean-sample variance of the frequency deviation

$$E(\hat{\sigma}_N^2) = \frac{2Jw_v}{T^2} \left[\Delta_v(T) - \frac{\Delta_v(NT)}{N^2} \right] \quad (38)$$

which for large N is the same as σ_v^2 .

G. On S/N Estimation, J. W. Layland

1. Introduction

The bit S/N estimator has been analyzed in general by Gilchrist (SPS 37-27, Vol. IV, p. 169) for the strong signal case. (If $\{X_i\}$ is the data sequence upon which the S/N estimate is to be based, strong signal is defined by $E\{|X_i|\} \approx |E\{X_i\}|$, where $E\{\}$ denotes expectation.) Two questions not answered by this analysis form the subject of the present article: (1) How does the S/N estimator behave with weak signals? (2) Is the S/N estimator a feasible in-lock indicator for a bit-synchronization loop?

2. The S/N Estimator in the Weak Signal Case

The estimator in question is constructed as follows: Let $y(t)$ be the received base-band signal, $\hat{\tau}_i$ the estimated data transition time at the start of the i th bit, and I_i the i th data bit integral. Then

$$I_i = \int_{\hat{\tau}_i}^{\hat{\tau}_{i+1}} y(t) dt \quad (1)$$

and the S/N estimate at the m th bit is constructed using the N -most-recent data bit integrals:

$$\hat{R} = \frac{\frac{1}{2} \left(\frac{1}{N} \sum_{i=1}^N |I_{m+1-i}| \right)^2}{\frac{1}{N-1} \sum_{i=1}^N \left(|I_{m+1-i}| - \frac{1}{N} \sum_{j=1}^N |I_{m+1-j}| \right)^2} \quad (2)$$

As N approaches infinity, the sample mean in the numerator of Eq. (2) converges (in probability) to $E\{|I|\}$, and the sample variance, the denominator of Eq. (2), converges to $V\{|I|\}$. For large, but finite N , the behavior of \hat{R} can be expressed in terms of perturbations about these asymptotic values. Define p_n^N and p_d^N such that

$$\hat{R} = \frac{1}{2} \frac{(E\{|I|\})^2}{V\{|I|\}} \left(1 + \frac{p_n^N}{E\{|I|\}} \right)^2 \left(1 + \frac{p_d^N}{V\{|I|\}} \right)^{-1} \quad (3)$$

$$\hat{R} = R^*(R) \left(1 + \frac{p_n^N}{E\{|I|\}} \right)^2 \left(1 + \frac{p_d^N}{V\{|I|\}} \right)^{-1}$$

where the notation $R^*(R)$ has been adopted for convenience, R being the true S/N (signal-power · bit-time/noise-spectral-density). The function $R^*(R)$ is plotted in

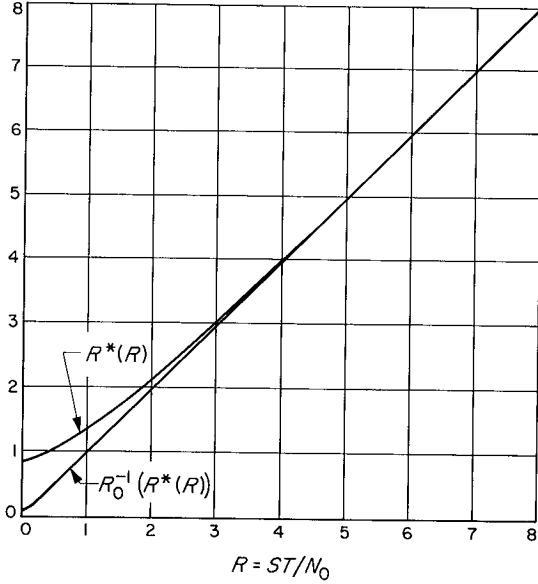


Fig. 12. Asymptotic values of the estimators \hat{R} and \hat{R}_0' vs R

Fig. 12. The perturbation terms both have zero mean and variances

$$V \left\{ \frac{p_n^N}{E\{|I|\}} \right\} = \frac{1}{N} \frac{1}{2R^*(R)}$$

$$V \left\{ \frac{p_d^N}{V\{|I|\}} \right\} = \frac{1}{N} \left[2 \left(\frac{1+2R^*}{1+2R} \right)^2 + 8 \left(\frac{R^*}{\pi} \right)^{1/2} e^{-R} \left(\frac{1+2R^*}{1+2R} \right)^{3/2} + 8(R-R^*) \left(\frac{1+2R^*}{1+2R} \right) \right] \quad (4)$$

For N large, all higher-order terms in the power-series expansion of \hat{R} may be dropped and

$$\hat{R} \approx R^*(R) [1 + \delta]$$

$$E\{\delta\} \approx 0$$

$$V\{\delta\} \approx \frac{1}{N} \left\{ \frac{2}{R^*} + 2 \left(\frac{1+2R^*}{1+2R} \right)^2 + 8 \left(\frac{R^*}{\pi} \right)^{1/2} e^{-R} \left(\frac{1+2R^*}{1+2R} \right)^{3/2} + 8(R-R^*) \left(\frac{1+2R^*}{1+2R} \right) \right\} \quad (5)$$

approximately.

The estimator \hat{R} is a biased estimator. Its large-sample asymptote $R^*(R)$ is given by

$$R^*(R) = \frac{1}{2} \frac{[E\{|I|\}]^2}{V\{|I|\}}$$

$$= \frac{\left[(R)^{1/2} (1 - 2 \operatorname{erfc}(2R)^{1/2}) + \frac{1}{(\pi)^{1/2}} e^{-R} \right]^2}{1 + 2 \left\{ R - \left[(R)^{1/2} (1 - 2 \operatorname{erfc}(2R)^{1/2}) + \frac{1}{(\pi)^{1/2}} e^{-R} \right]^2 \right\}} \quad (6)$$

Since $R^*(R)$ is a function of R , it possesses a unique inverse; call it $R^{-1}(x)$. The estimator \hat{R}' defined by $\hat{R}' = R^{-1}(\hat{R})$ is an asymptotically unbiased estimator of R for all values of R .

The behavior of \hat{R}' for large N may be seen as follows: expand $R^{-1}(x)$ in a Taylor series about $R^*(R)$.

$$\hat{R}' = R^{-1}(R^*) + (\hat{R} - R^*) \left. \frac{d}{dx} R^{-1}(x) \right|_{x=R^*} + \text{higher-order terms} \quad (7)$$

If N is large enough, the convergence of the perturbation terms in Eq. (3) implies that the higher-order terms in Eq. (7) may be neglected and

$$E\{\hat{R}'\} \cong R$$

$$V\{\hat{R}'\} \cong V\{\hat{R}\} \cdot \left[\left. \frac{d}{dx} R^{-1}(x) \right|_{x=R^*(R)} \right]^2 \quad (8)$$

$$\cong V\{\hat{R}\} \cdot \left[\left. \frac{d}{dx} R^*(x) \right|_{x=R} \right]^{-2}$$

And if N is large enough, both Eqs. (5) and (8) will be valid, and

$$V\{\hat{R}'\} \approx \frac{1}{N} \left\{ 2R^* + R^{*2} \left[2 \left(\frac{1+2R^*}{1+2R} \right)^2 + 8 \left(\frac{R^*}{\pi} \right)^{1/2} e^{-R} \left(\frac{1+2R^*}{1+2R} \right)^{3/2} + 8(R-R^*) \left(\frac{1+2R^*}{1+2R} \right) \right] \left[\left. \frac{d}{dx} R^*(x) \right|_{x=R} \right]^{-2} \right\} \quad (9)$$

The conditions under which Eqs. (5) and (8) are valid are, essentially, $V\{\hat{R}\} \ll R^{*2}$ and $V\{\hat{R}\} \ll 1$. If reasonably tight requirements are placed upon the behavior of \hat{R}' , the validity conditions will, in most circumstances, follow as a side effect of these requirements, and Eq. (9) may then be used to determine the minimal N which will satisfy the assigned conditions.

Computation of $R^{-1}(x)$ is easily accomplished by means of polynomial approximation. The relationship of $(\pi - 2)(R^* - R)$ to $1/(\pi - 2)R^*$, plotted in Fig. 13, has been fitted with a piecewise second-order least-squares approximation which is correct to within 0.01 for $R > 0.1$. This approximation, call it $R_0^{-1}(x)$, admits the following description: If $x \geq 6.91$, $R_0^{-1}(x) = x$, or if $6.91 > x \geq 3.04$,

$$R_0^{-1}(x) = x - (\pi - 2)^{-1} (0.156 + (\pi - 2)^{-1} x^{-1} (-0.409 + 2.412 (\pi - 2)^{-1} x^{-1}))$$

or if $3.04 > x \geq 1.08$,

$$R_0^{-1}(x) = x - (\pi - 2)^{-1} (-0.194 + (\pi - 2)^{-1} x^{-1} (0.9963 + 0.0257 (\pi - 2)^{-1} x^{-1})) \quad (11)$$

or if $1.08 > x$,

$$R_0^{-1}(x) = x - (\pi - 2)^{-1} (0.8103 + (\pi - 2)^{-1} x^{-1} (-1.501 + 1.581 (\pi - 2)^{-1} x^{-1}))$$

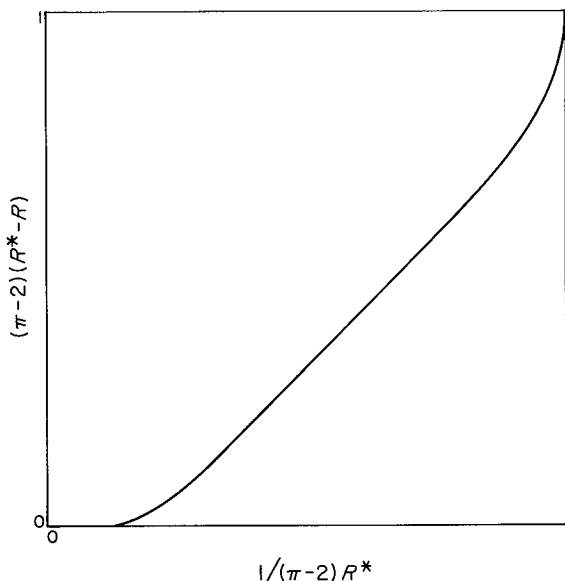


Fig. 13. Error term in $R^*(R)$

The asymptotic value of the estimator \hat{R}' , using this particular inversion, i.e., $R_0^{-1}(R^*(R))$, is also plotted versus R in Fig. 12. The S/N estimator of the multiple-mission-telemetry demonstration system is \hat{R}'_0 . Fig. 14 shows a set of experimental cdf's for this estimator for $N = 24$, $R = 0.5, 1, 2, 4$. Experimental cdf's obtained for higher values of R are virtually identical to that for $R = 4$.

3. S/N Estimator as a Lock Detector

The relationship between the timing references and I_i , and hence (in qualitative terms) \hat{R} , is easily determined:

$$I_i = \int_{\hat{\tau}_i}^{\hat{\tau}_{i+1}} n(t) dt + \int_{\hat{\tau}_i}^{\tau_i} m_{i-1} A \cos \theta_{rf} \left(1 - \frac{2}{\pi} \left| \theta_{sc} \right| \right) dt + \int_{\tau_i}^{\hat{\tau}_{i+1}} m_i A \cos \theta_{rf} \left(1 - \frac{2}{\pi} \left| \theta_{sc} \right| \right) dt \quad (12)$$

Where $y(t) = n(t) + A m(t) \cos \theta_{rf} (1 - 2/\pi |\theta_{sc}|)$ is the low-frequency component of the demodulated received signal, θ_{rf} and θ_{sc} are the RF and subcarrier phase errors, assumed constant over T_B , and τ_i is the true start of the i th bit, assumed greater than $\hat{\tau}_i$. Let $\theta = (\tau_i - \hat{\tau}_i)$. Then if the integral crosses a transition, the mean of I_i is $m_i AT_B \cos \theta_{rf} (1 - (2/\pi) |\theta_{sc}|) (1 - 2|\theta|)$; while if it crosses no transition, the mean is $m_i AT_B \cos \theta_{rf} (1 - (2/\pi) |\theta_{sc}|)$.

The estimate is constructed of samples from populations with these two differing means, and the distribution of samples between the population is, unfortunately, data-dependent. Assume that data transitions occur with probability $1/2$. Then it can be shown that if R is large, and N approaches infinity, \hat{R} converges to

$$\hat{R} \rightarrow R_0 = \frac{R_1 (1 - |\theta|)^2}{1 + 2R_1 \theta^2}, \quad |\theta| \leq 1/2 \quad (13)$$

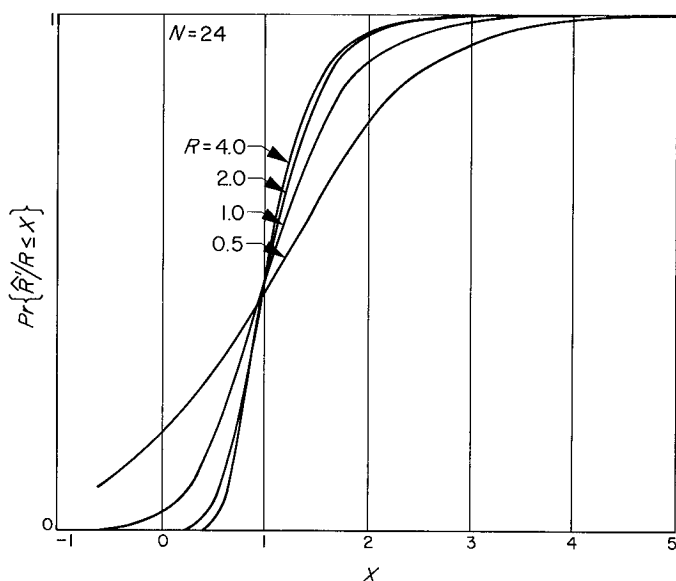


Fig. 14. The cdf of \hat{R}/R for $R = 0.5, 1, 2, 4$ ($N = 24$)

if θ is essentially fixed over the measurement period, and

$$R_1 = \frac{ST_B}{N_0} \overline{\cos^2 \theta} f \left(1 - \frac{2}{\pi} \theta_{sc} \right)^2$$

Another situation of interest is that of slipping cycles. If several cycles are slipped slowly over a long measurement period, then the signal amplitude, and signal-plus-noise power estimates which are components of \hat{R} are approximately averaged over all values of θ . In this case, \hat{R} converges to

$$\hat{R} \rightarrow R_s = \frac{27R_1}{48 + 10R_1} \quad (14)$$

The residual signal term in the denominators of Eqs. (13) and (14) is clearly beneficial to the identification of the out-of-lock condition.

While the asymptotic values, R_o and R_s , may suggest the lock-detector's performance, they are an incomplete answer because the measurement time N , being proportional to the measurement time of the bit-time-tracking filter, is, generally, insufficient to assure the convergence of the S/N estimator. The distribution of the estimator must be determined to completely evaluate this lock detector. For the strong-signal case, where the distributions of the sample mean and variance are known, an

analytic closed-form solution should be attainable, but is not yet accomplished. Experimental cdf's have been determined for various values of R and θ . Two families of these are shown in Figs. 15 and 16. From these curves, it appears as if the desired lock-detector performance can be obtained from the S/N estimator.

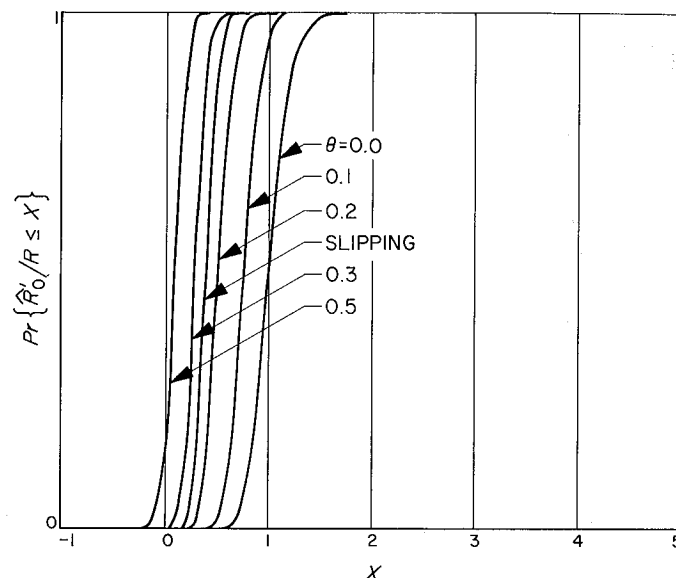


Fig. 15. Experimental cdf for $R = 3.16$, $N = 100$, and various θ

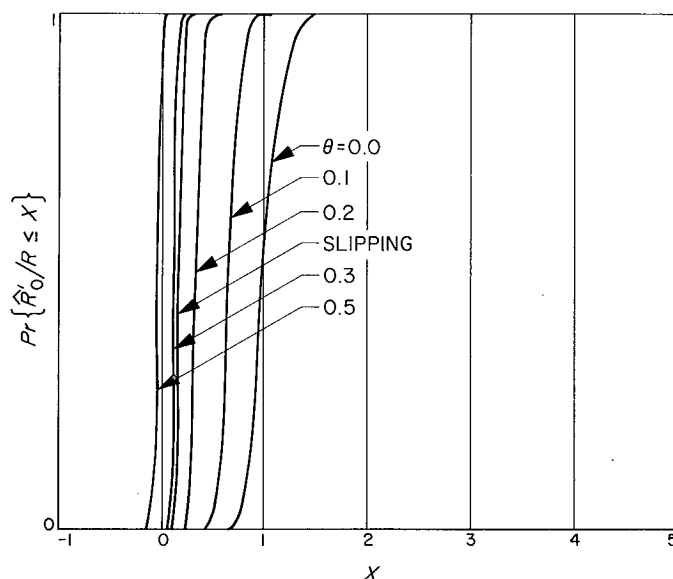


Fig. 16. Experimental cdf for $R = 10$, $N = 100$, and various θ

H. Digital Filtering of Random Sequences,

G. Jennings

1. Introduction

The Deep Space Network is beginning to rely on digital techniques to perform functions formerly performed by analog elements of telemetry receivers. Some of the reasons for this trend are the non-existence of stable analog elements with long time constants, the need to save labor at tracking stations by running most of the operation by computer, and the need to save money by outfitting the stations with mission-independent equipment. To achieve these goals, certain analog elements, for example, phase-locked loops, can be replaced by digital circuits or computer programs. This article studies the effect of quantization and round-off error on such systems. In particular, we study the case of digital filtering. Thus, consider filtering a random sequence $\{x_1, x_2, \dots\}$ to form another sequence $\{y_1, y_2, \dots\}$ by a linear recurrence of the form

$$y_n = \sum_{k=1}^K a_k y_{n-k} + x_n \quad (1)$$

In some applications, such as when the random variables lie in a finite field, the $\{y_n\}$ can be computed exactly. If the values of the $\{x_n\}$ are arbitrary real numbers, however, as in the case of signals plus Gaussian noise, Eq. (1) can only be solved approximately. The sequence $\{y_n\}$ which is found actually satisfies

$$y_n = \sum_{k=1}^K a_k y_{n-k} + x_n + \delta_n \quad (2)$$

where δ_n is the error in evaluating the right side of Eq. (1).

The error we consider is of the type which occurs if we try to solve Eq. (1) on a digital computer. Suppose that fixed-point numbers are used to store the values of the $\{y_n\}$. Then there is a finite sequence of possible values of these variables. This set of values P has the form

$$P : \{0, \pm 2\delta, \pm 4\delta, \dots, \pm 2k_0\delta = \pm M\}$$

where we denote the separation between levels of 2δ for convenience. We suppose that δ_n is chosen to be the number of smallest absolute value so that the value of y_n given by Eq. (2) lies in the set P . Many of the estimates to be derived can be extended to a more general case, in which we assume only that δ_n is of order δ when

the quantity

$$\sum_{k=1}^K a_k y_{n-k} + x_n$$

is not outside the range spanned by P .

The sequence $\{x_1, x_2, \dots\}$ is taken to be a stationary Gaussian process of mean zero, variance σ^2 . To distinguish between the solutions of Eqs. (1) and (2), we will denote the solution of Eq. (1) by $(\tilde{y}_1, \tilde{y}_2, \dots)$, rewriting that equation as

$$\tilde{y}_n = \sum_{k=1}^K a_k \tilde{y}_{n-k} + x_n \quad (1')$$

We want to compare the solution of Eq. (2) with the solution of Eq. (1') when the same initial values are used for each sequence:

$$y_n = \tilde{y}_n = y_n^{(0)}, \quad n = 0, 1, \dots, K-1$$

It is assumed that the solution of Eq. (1') is stable in the sense that the values of y_n when n is large are affected very little by the initial values $y_0^{(0)}, \dots, y_{K-1}^{(0)}$. This is true if all the roots of the polynomial

$$t^K - \sum_{k=1}^K a_k t^{K-k} \quad (3)$$

have absolute values less than 1. *A priori*, it is not clear whether this forces stability of the solution of Eq. (2).

With this assumption on the $\{a_k\}$, the solution of Eq. (1') approaches a stationary Gaussian process for n large. Quantities of interest for this process are the covariances $E(\tilde{y}_n \tilde{y}_m)$ and the spectrum of the limiting stationary process. To determine by how much estimates for these quantities can be affected when we use the solution of Eq. (2) instead of Eq. (1'), it is sufficient to estimate

$$E[(y_n - \tilde{y}_n)(y_m - \tilde{y}_m)]$$

A bound which approaches zero as $|m - n| \rightarrow \infty$ is needed to estimate the effect of the errors on the spectrum.

In Part 2, a special case of Eq. (2) is treated in detail. Following this, we deal with a fairly general case, in which the only assumption made is that

$$\sum_{k=1}^K |a_k| < 1$$

This is a greater restriction than the condition that the roots of Eq. (3) be less than 1. It is used to control the behavior of the sum on the right in Eq. (2). The methods developed for the first-order recurrence are generalized to this case. The general stationary Gaussian process $\{x_n\}$ is approximated by a finite moving average of a sequence of independent random variables (Eq. 15). Because of the similarity to the first-order case, the development is not carried out in full detail in Part 3.

2. A Special Case

Here we assume that the recurrence relation is of order 1 and the x_i 's are independent. The basic equation is

$$\tilde{y}_i = \lambda \tilde{y}_{i-1} + x_i \quad (4)$$

and the approximating equation is

$$y_i = \lambda y_{i-1} + x_i + \delta_i \quad (5)$$

It is assumed that $|\lambda| < 1$.

The solutions of Eqs. (4) and (5) can be written explicitly. They are

$$\begin{aligned} \tilde{y}_i &= \lambda^i y_0 + \sum_{j=1}^i \lambda^{i-j} x_j \\ y_i &= \lambda^i y_0 + \sum_{j=1}^i \lambda^{i-j} (x_j + \delta_j) \end{aligned}$$

Hence

$$y_i - \tilde{y}_i = \sum_{j=1}^i \lambda^{i-j} \delta_j \quad (6)$$

It follows that

$$\int (y_i - \tilde{y}_i)(y_k - \tilde{y}_k) d\mu = \sum_{j=1}^i \sum_{l=1}^k \lambda^{i+k-j-l} \int \delta_j \delta_l d\mu \quad (7)$$

Hence,

$$\begin{aligned} \int_{S_{i-1}(n_j) \times (M - \lambda n_j - \delta, \infty)} \delta_i^2 d\mu_i &= \int_{S_{i-1}(n_j)} d\mu_{i-1} \int_{M - \lambda n_j - \delta}^{\infty} [x - (M - \lambda n_j)]^2 \frac{\exp\left(-\frac{x^2}{2\sigma^2}\right)}{(2\pi)^{1/2} \sigma} dx \\ &\leq \delta^2 \mu[S_{i-1}(n_j) \times (M - \lambda n_j - \delta, \infty)] + \int_{S_{i-1}(n_j)} d\mu_{i-1} \int_{M - \lambda n_j}^{\infty} [x - (M - \lambda n_j)]^2 \frac{\exp\left(-\frac{x^2}{2\sigma^2}\right)}{(2\pi)^{1/2} \sigma} dx \end{aligned} \quad (10)$$

where μ denotes probability measure. We proceed to estimate the integrals on the right in this equation. First the case of equal subscripts is treated.

Lemma 1. If $\{y_i\}, \{\delta_i\}$ satisfy Eq. (5) with $|\lambda| < 1$, where the x_i are independent normal with mean 0, variance σ^2 , then

$$\int \delta_i^2 d\mu < \delta^2 + \frac{2\sigma^2}{(2\pi)^{1/2}} \frac{1}{M(1-|\lambda|)} \exp\left[-\frac{M^2(1-|\lambda|)^2}{2\sigma^2}\right] \quad (8)$$

Proof. For fixed i , and $n_k \in P$, we make the following definition:

$$S_i(n_k) = \{(x_1, \dots, x_i) : y_i = n_k\}$$

Clearly, R^i is the disjoint union over all k of $S_i(n_k)$. If $y_i = n_k$, where n_k is not equal to $\pm M$, then δ_i must be less or equal to δ in magnitude. Hence,

$$\begin{aligned} \int \delta_i^2 d\mu &\leq \delta^2 \{1 - \mu[S_i(M)] - \mu[S_i(-M)]\} \\ &\quad + \int_{S_i(M)} \delta_i^2 d\mu + \int_{S_i(-M)} \delta_i^2 d\mu \end{aligned} \quad (9)$$

We need therefore only estimate the last two integrals. Note that

$$S_i(M) = \sum_j S_{i-1}(n_j) \times (M - \lambda n_j - \delta, \infty)$$

where Σ denotes the union of disjoint sets. Let

$$(x_1, \dots, x_i) \in S_{i-1}(n_j) \times (M - \lambda n_j - \delta, \infty)$$

Then

$$\delta_i(x_1, \dots, x_i) = -[x_i - (M - \lambda n_j)]$$

We note that for $z > 0$,

$$\int_z^\infty (t-z)^2 \exp\left(-\frac{t^2}{2\sigma^2}\right) dt < \frac{\sigma^4}{z} \exp\left(-\frac{z^2}{2\sigma^2}\right) \quad (11)$$

We have required that $|\lambda| < 1$, and hence obtain that $M - \lambda n_j > (1 - |\lambda|)M > 0$. Summing Eq. (10) over all values of j and using the estimate of Eq. (11) we obtain that

$$\int_{S_i(M)} \delta_i^2 d\mu < \frac{\sigma^3}{(2\pi)^{1/2}(1-|\lambda|)M} \exp\left[-\frac{(1-|\lambda|)^2 M^2}{2\sigma^2}\right] + \delta^2 \mu[S_i(M)]$$

A similar result obtains for

$$\int_{S_i(-M)} \delta_i^2 d\mu$$

Using this bound in Eq. (9), the lemma is proved.

Lemma 2. Let A be a measurable subset of R^i [the space of points (x_1, \dots, x_i)] of the form

$$A_{i-1} \times \{-\infty < x_i < \infty\}$$

Then, under the hypotheses of lemma 1,

$$\int_A |\delta_i(x_1, \dots, x_i)| d\mu \leq \mu(A_{i-1}) \times \left[\frac{2\sigma}{(2\pi)^{1/2}} \exp\left(-\frac{M^2(1-|\lambda|)^2}{2\sigma^2}\right) + \delta \right]$$

Proof. We note that

$$A = \sum_j A \cap S_i(n_j)$$

As before,

$$\begin{aligned} \int_A |\delta_i(x_1, \dots, x_i)| d\mu &\leq \delta \{ \mu(A) - \mu[S_i(M) \cap A] \\ &\quad - \mu[S_i(-M) \cap A] \} \\ &\quad + \int_{S_i(M) \cap A} |\delta_i| d\mu \\ &\quad + \int_{S_i(-M) \cap A} |\delta_i| d\mu \quad (12) \end{aligned}$$

Therefore, we need only estimate each of the last two integrals. To do this, we note the following decomposition:

$$\begin{aligned} S_i(M) \cap A &= \sum_j [S_{i-1}(n_j) \times (M - \lambda n_j - \delta, \infty)] \cap A \\ &= \sum_j [S_{i-1}(n_j) \cap A_{i-1}] \times (M - \lambda n_j - \delta, \infty) \end{aligned}$$

We shall integrate over each of the sets in the last sum.

$$\begin{aligned} \int_{[S_{i-1}(n_j) \cap A_{i-1}] \times (M - \lambda n_j - \delta, \infty)} |\delta_i| d\mu &= \int_{S_{i-1}(n_j) \cap A_{i-1}} d\mu \int_{M - \lambda n_j - \delta}^\infty |x - (M - \lambda n_j)| \frac{\exp\left(-\frac{x^2}{2\sigma^2}\right)}{(2\pi)^{1/2} \sigma} dx \\ &\leq \mu[S_{i-1}(n_j) \cap A_{i-1}] \left[\delta + \int_{M - \lambda n_j}^\infty \frac{x \exp\left(-\frac{x^2}{2\sigma^2}\right)}{(2\pi)^{1/2} \sigma} dx \right] \\ &= \mu[S_{i-1}(n_j) \cap A_{i-1}] \left[\delta + \frac{\sigma}{(2\pi)^{1/2}} \exp\left(-\frac{(M - \lambda n_j)^2}{2\sigma^2}\right) \right] \end{aligned}$$

Now summing the last inequality over j , and making the estimate $M - \lambda n_j \geq (1 - |\lambda|)M$, we conclude

$$\begin{aligned} \int_{S_i(M) \cap A} |\delta_i| d\mu &\leq \mu[S_i(M) \cap A_{i-1}] \\ &\quad \times \left[\delta + \frac{\sigma}{(2\pi)^{1/2}} \exp\left(-\frac{M^2(1-|\lambda|)^2}{2\sigma^2}\right) \right] \end{aligned}$$

Similarly,

$$\begin{aligned} \int_{S_i(-M) \cap A} |\delta_i| d\mu &\leq \mu[S_i(-M) \cap A_{i-1}] \\ &\quad \times \left[\delta + \frac{\sigma}{(2\pi)^{1/2}} \exp\left(-\frac{M^2(1-|\lambda|)^2}{2\sigma^2}\right) \right] \end{aligned}$$

Putting these last two estimates into Eq. (12), we obtain the result of the lemma.

Lemma 3. Let

$$P = \frac{2}{(2\pi)^{1/2} \sigma} \int_0^{(1+|\lambda|)M} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx$$

Then, under the hypotheses of lemma 1, for $j > k$

$$|f \delta_j \delta_k d\mu| \leq \left(\frac{\sigma}{(2\pi)^{1/2}} \exp\left[-\frac{M^2(1-|\lambda|)^2}{2\sigma^2}\right] + \delta \right)^2 P^{j-k-1}$$

Proof. We have

$$\begin{aligned} |f \delta_j \delta_k d\mu| &= |f \delta_k (f \delta_j d\mu_{k+1} \cdots d\mu_j) d\mu_1 \cdots d\mu_k| \\ &\leq f |\delta_k| |f \delta_j d\mu_{k+1} \cdots d\mu_j| d\mu_1 \cdots d\mu_k \end{aligned} \quad (13)$$

Let $R_{k,j} = \{(x_{k+1}, \dots, x_j) : |x_s| \leq (1+|\lambda|)M \text{ for } k < s < j\}$. If (x_{k+1}, \dots, x_j) lies in the complement $R'_{k,j}$ of $R_{k,j}$, δ_j is independent of (x_1, \dots, x_k) and y_0 , for if

$$\pm x_s > (1+|\lambda|)M, \quad y_s = \pm M$$

δ_j is clearly an odd function of x_1, \dots, x_j and y_0 :

$$\delta_j(x_1, \dots, x_j; y_0) = -\delta_j(-x_1, \dots, -x_j; -y_0)$$

from the form of Eq. (5) and the symmetry of the set N . Hence, in $R'_{k,j}$

$$\begin{aligned} \delta_j(x_1, \dots, x_j; y_0) &= \delta_j(-x_1, \dots, -x_k, x_{k+1}, \dots, x_j; -y_0) \\ &= -\delta_j(x_1, \dots, x_k, -x_{k+1}, \dots, -x_j; y_0) \end{aligned}$$

Since the region $R'_{k,j}$ is symmetric about 0,

$$\int_{R'_{k,j}} \delta_j d\mu_{k+1} \cdots d\mu_j = 0$$

This reduces Eq. (13) to

$$|f \delta_j \delta_k d\mu| \leq f |\delta_k| |f_{R_{k,j}} \delta_j d\mu_{k+1} \cdots d\mu_j| d\mu_1 \cdots d\mu_k \quad (14)$$

The inner integral here is a function of x_1, \dots, x_k to which lemma 2 can be applied, since

$$\delta_j(x_1, \dots, x_j; y_0) = \delta_{j-k}(x_{k+1}, \dots, x_j; y_k)$$

We have $\mu(R_{j,j}) = P^{j-k-1}$, and

$$\begin{aligned} |f_{R_{k,j}} \delta_j d\mu_{k+1} \cdots d\mu_j| &\leq P^{j-k-1} \\ &\times \left\{ \frac{\sigma}{(2\pi)^{1/2}} \exp\left[-\frac{M^2(1-|\lambda|)^2}{2\sigma^2}\right] + \delta \right\} \end{aligned}$$

Hence, Eq. (14) implies

$$\begin{aligned} |f \delta_j \delta_k d\mu| &\leq P^{j-k-1} \\ &\times \left\{ \frac{\sigma}{(2\pi)^{1/2}} \exp\left[-\frac{M^2(1-|\lambda|)^2}{2\sigma^2}\right] + \delta \right\} f |\delta_k| d\mu \end{aligned}$$

and another application of lemma 2 completes the proof.

Lemma 4. Under the same hypotheses as in lemma 3,

$$\begin{aligned} |f \delta_j \delta_k d\mu| &\leq \\ &\left\{ \frac{2\sigma^3}{(2\pi)^{1/2} M(1-|\lambda|)} \exp\left[-\frac{M^2(1-|\lambda|)^2}{2\sigma^2}\right] + \delta^2 \right\} P^{(j-k-1)/2} \end{aligned}$$

Proof. By Schwarz's inequality,

$$|f_{R_{k,j}} \delta_j d\mu_{k+1} \cdots d\mu_j|^2 \leq \mu(R_{k,j}) f \delta_j^2 d\mu_{k+1} \cdots d\mu_j$$

Applying Schwarz's inequality also to Eq. (14), we have

$$\begin{aligned} |f \delta_j \delta_k d\mu|^2 &\leq f \delta_k^2 d\mu f |f_{R_{k,j}} \delta_j d\mu_{k+1} \cdots d\mu_j|^2 d\mu_1 \cdots d\mu_k \\ &\leq f \delta_k^2 d\mu \cdot \mu(R_{k,j}) f \delta_j^2 d\mu \end{aligned}$$

Estimating each of these integrals by lemma 1 proves lemma 4.

We now use all of these estimates together with Eq. (7) to estimate $f(y_i - \tilde{y}_i)^2$.

Theorem 1. Let $|\lambda| < 1$, and as before let

$$y_i = \lambda y_{i-1} + x_i + \delta_i$$

$$y_0 = y_0$$

Further define

$$\tilde{y}_i = \lambda \tilde{y}_{i-1} + x_i$$

$$\tilde{y}_0 = y_0$$

The x_i 's and δ_i 's are as before. Under these conditions

$$\begin{aligned} f(\tilde{y}_i - y_i)^2 d\mu \leq & \frac{1}{1 - \lambda^2} \left\{ \delta^2 + \frac{2\sigma^3}{(2\pi)^{1/2}} \frac{1}{M(1 - |\lambda|)} \exp \left[-\frac{M^2(1 - |\lambda|)^2}{2\sigma^2} \right] \right\} \\ & + \frac{2|\lambda|}{(1 - \lambda^2)(1 - P|\lambda|)} \left\{ \delta + \frac{2\sigma}{(2\pi)^{1/2}} \exp \left[-\frac{M^2(1 - |\lambda|)^2}{2\sigma^2} \right] \right\}^2 \end{aligned}$$

where

$$P = 2 \int_0^{M(1+|\lambda|)} \frac{\exp\left(-\frac{x^2}{2\sigma^2}\right)}{(2\pi)^{1/2}\sigma} dx$$

Proof. Equation (7) yields

$$\begin{aligned} \int (y_i - \tilde{y}_i)^2 d\mu &= \sum_{j,k=1}^i \lambda^{2i-j-k} \int \delta_j \delta_k d\mu \\ &= \sum_{j=1}^i \lambda^{2i-2j} \int \delta_j^2 d\mu \\ &\quad + 2 \sum_{i \geq j > k \geq 1} \lambda^{2i-j-k} \int \delta_j \delta_k d\mu \end{aligned}$$

Lemmas 1 and 3 yield the following upper bound:

$$\begin{aligned} & \left\{ \delta^2 + \frac{2\sigma^3}{(2\pi)^{1/2}} \frac{1}{M(1 - |\lambda|)} \exp \left[-\frac{M^2(1 - |\lambda|)^2}{2\sigma^2} \right] \right\} \sum_{j=1}^i |\lambda|^{2(i-j)} \\ & + 2 \left\{ \delta + \frac{2\sigma}{(2\pi)^{1/2}} \exp \left[-\frac{M^2(1 - |\lambda|)^2}{2\sigma^2} \right] \right\}^2 \sum_{i \geq j > k \geq 1} |\lambda|^{2i-j-k} P^{j-k-1} \end{aligned}$$

Now

$$\begin{aligned} \sum_{i \geq j > k \geq 1} |\lambda|^{2i-j-k} P^{j-k-1} &= \sum_{j=2}^i |\lambda|^{2i-2j} \sum_{k=1}^{j-1} (|\lambda|P)^{j-k} \\ &\leq \sum_{j=2}^i |\lambda|^{2i-2j} \frac{|\lambda|P}{1 - |\lambda|P} \\ &\leq \frac{1}{1 - \lambda^2} \frac{|\lambda|P}{1 - |\lambda|P} \end{aligned}$$

This latter estimate together with the fact that

$$\sum_{j=1}^i \lambda^{2i-2j} \leq \frac{1}{1 - \lambda^2}$$

proves the theorem.

Theorem 2. Under the same conditions as theorem 1,

$$\int (y_i - \tilde{y}_i)^2 d\mu \leq \frac{1}{1 - \lambda^2} \left[1 + \frac{2|\lambda|}{1 - |\lambda|(P)^{1/2}} \right] \\ \times \left\{ \delta^2 + \frac{2\sigma^3}{(2\pi)^{1/2}} \frac{1}{M(1 - |\lambda|)} \exp \left[-\frac{M^2(1 - |\lambda|)^2}{2\sigma^2} \right] \right\}$$

Proof. The proof proceeds exactly as in the proof of theorem 1, but the estimate of lemma 4 is used instead of that of lemma 3.

Lemmas 3 and 4 can be applied in the same way to Eq. (7) when $i \neq k$. For example, lemma 4 implies the following:

Theorem 3. Under the hypotheses of theorem 1, for $i > k$

$$|f(y_i - \tilde{y}_i)(y_k - \tilde{y}_k) d\mu| \leq \left\{ \frac{|\lambda|^{i-k}}{1 - \lambda^2} \left[1 + \frac{2|\lambda|}{1 - |\lambda|(P)^{1/2}} \right] \right. \\ \left. + \frac{P^{(i-k)/2} - |\lambda|^{i-k}}{(P)^{1/2} - |\lambda|} \right\} \\ \times \left\{ \delta^2 + \frac{2\sigma^3}{(2\pi)^{1/2}} \frac{1}{M(1 - |\lambda|)} \exp \left[-\frac{M^2(1 - |\lambda|)^2}{2\sigma^2} \right] \right\}$$

3. Equations of Arbitrary Order

Here we consider the general equation

$$y_i = \sum_{k=1}^K a_k y_{i-k} + \sum_{j=0}^L b_j x_{i-j} + \delta_i \quad (15)$$

with y_0, \dots, y_{K-1} given. We make the assumption

$$a = \sum_{k=1}^K |a_k| < 1$$

which makes the methods of Part 2 applicable.

We begin by stating without proof two lemmas needed for the representation of the solution of Eq. (15).

Lemma 5. The sequences which satisfy the linear recurrence

$$y_i = a_1 y_{i-1} + a_2 y_{i-2} + \dots + a_K y_{i-K} \quad (16)$$

form a linear space of dimension K . Moreover, let

$$x^K - a_1 x^{K-1} - a_2 x^{K-2} - \dots - a_K = \prod_{i=1}^j (x - \lambda_i)^{e_i}$$

then the space of solutions of the linear recurrence Eq. (16) is spanned by the following set of sequences:

$$\{\lambda_i^r\}_{r=0}^{\infty}$$

together with for each $e_i > 1$, the class of sequences $\{(r+1)(r+2)\dots(r+t)\lambda_i^r\}$ for $1 \leq t < e_i$, and $1 \leq i \leq j$.

Before continuing, we make the definition of the matrix A to be the $K \times K$ matrix which has the following k -vectors as columns:

$$\{\lambda_i^r\}_{r=0}^{K-1}$$

and if $1 < e_i$, also the vectors

$$\{(r+1)(r+2)\dots(r+t)\lambda_i^r\}_{r=0}^{K-1}$$

for $1 \leq t < e_i$ and $1 < i \leq j$. Note that the columns of A are linearly independent and hence A is invertible. Also define a K -vector $\Phi(s)$ depending on s which has as components the value of one of the following sequences:

$$\{\lambda_i^s\}, \{(s+1)\dots(s+t)\lambda_i^s\}$$

for $1 \leq t < e_i$ and $1 \leq i \leq j$, in the same order as the corresponding vector appears as a column of A . We will now state a formula for the solution of inhomogeneous linear recurrence in terms of the solution to the homogeneous equation.

Lemma 6. The linear recurrence with inhomogeneous terms

$$y_i = a_1 y_{i-1} + \dots + a_K y_{i-K} + x_i \quad (17)$$

subject to

$$y_0 = y_1 = \dots = y_{K-1} = 0$$

has a unique solution which is given for i greater than or equal to by

$$y_i = \sum_{j=K}^i \{A^{-1} e_K \cdot \Phi(K-1+i-j)\} x_j \quad (18)$$

where e_K is the column vector consisting of zeros except for a one in the last position.

Proof. The existence of an unique solution is immediate. Moreover, that the Formula (18) satisfies the recurrence Eq. (17) is a straightforward computation.

Define as in Part 1

$$S_i(M) = \{(x_{K-L}, \dots, x_i) : y_i(x_{K-L}, \dots, x_i) = M\}$$

and

$$S_i(-M) = \{(x_{K-L}, \dots, x_i) : y_i(x_{K-L}, \dots, x_i) = -M\}$$

Hence, if $(x_{K-L}, \dots, x_i) \notin S_i(+M) \cup S_i(-M)$, then $\delta_i(x_{K-L}, \dots, x_i)$ is less than or equal to δ in magnitude.

Hence, we obtain

$$\int \delta_i^2 d\mu \leq \delta^2 \{1 - \mu[S_i(M) \cup S_i(-M)]\} + \int_{S_i(M) \cup S_i(-M)} \delta_i^2 d\mu \quad (19)$$

We only need to estimate the integral of δ_i^2 over $S_i(+M)$ and $S_i(-M)$. We shall deal only with the integral over the former. A similar result will obtain for the latter. On $S_i(+M)$ we notice that

$$-\delta_i = \sum_{j=0}^L b_j x_{i-j} - (M - \sum_{j=1}^K a_j y_{i-j})$$

We also observe the characterization of $S_i(+M)$ as the set such that

$$\sum_{j=1}^K a_j y_{i-j} + \sum_{j=0}^L b_j x_{i-j} \geq M - \delta$$

Hence, we obtain that

$$\begin{aligned} \int_{S_i(M)} \delta_i^2 d\mu &= \int_{\sum b_j x_{i-j} \geq M - \delta - \sum a_j y_{i-j}} [\sum b_j x_{i-j} - (M - \sum a_j y_{i-j})]^2 d\mu \\ &\leq \delta^2 + \int_{\sum b_j x_{i-j} \geq M - \delta - \sum a_j y_{i-j}} [\sum b_j x_{i-j} - (M - \sum a_j y_{i-j})]^2 d\mu \end{aligned} \quad (20)$$

Now we have required that $\sum |a_j| < 1$; hence, we have that

$$M - \sum a_j y_{i-j} \geq M(1 - \sum |a_j|) > 0$$

Moreover, we have

$$\sum b_j x_{i-j} - (M - \sum a_j y_{i-j}) \leq \sum b_j x_{i-j} - M(1 - \sum |a_j|)$$

Hence, we obtain by using these estimates in Relation (20) that

$$\begin{aligned} \int_{S_i(M)} \delta_i^2 d\mu &\leq \delta^2 \mu[S_i(M)] \\ &\quad + \int_{\sum b_j x_{i-j} \geq M(1-a)} [\sum b_j x_{i-j} - M(1-a)]^2 d\mu \end{aligned} \quad (21)$$

The random variable $\sum b_j x_{i-j}$ is normal, with mean zero, variance $\sigma^2 \sum b_j^2$. Thus, proceeding as in the proof of lemma 1, we get:

Lemma 7. Let y_i be the solution of the inhomogeneous linear recurrence

$$y_i = \sum_{j=1}^K a_j y_{i-j} + \sum_{j=0}^L b_j x_{i-j} + \delta_i$$

where the initial conditions are arbitrary, and the x_i 's are independent random variables with a Gaussian distribution of mean zero and standard deviation σ . The sequence of δ_i 's is chosen as detailed above. Moreover, assume that

$$a = \sum |a_j| < 1$$

Then

$$\int \delta_i^2 d\mu \leq \delta^2 + \frac{2\nu^3}{(2\pi)^{1/2} M(1-a)} \exp\left[-\frac{M^2(1-a)^2}{2\nu^2}\right]$$

where

$$\nu^2 = \sigma^2 \sum_{j=0}^L b_j^2$$

We are now ready to determine an estimate for the difference between the discrete problem of Eq. (15) and the continuous problem, where \tilde{y}_i is the solution to

$$\tilde{y}_i = \sum_{j=1}^K a_j \tilde{y}_{i-j} + \sum_{j=0}^L b_j x_{i-j} \quad (22)$$

with the same initial conditions as in Eq. (15). Upon subtracting Eq. (22) from (15) we obtain that

$$y_i - \tilde{y}_i = \sum_{j=1}^K a_j (y_{i-j} - \tilde{y}_{i-j}) + \delta_i$$

$$y_0 - \tilde{y}_0 = y_1 - \tilde{y}_1 = \dots = y_{K-1} - \tilde{y}_{K-1} = 0 \quad (23)$$

Application of lemma 6 yields

$$y_i - \tilde{y}_i = \sum_{j=K}^i [A^{-1} e_K \cdot \Phi(K-1+i-j)] \delta_i \quad (24)$$

Application of Schwarz's inequality yields the estimate that

$$[f(y_i - \tilde{y}_i)^2 d\mu]^{1/2}$$

$$\leq \sum_{j=K}^i |A^{-1} e_K \cdot \Phi(K-1+i-j)| (f\delta_j^2 d\mu)^{1/2}$$

$$\leq \sup_j f\delta_j^2 d\mu^{1/2} \sum_{j=K}^i |A^{-1} e_K \cdot \Phi(K-1+i-j)|$$

The sum

$$\sum_{j=K}^i |A^{-1} e_K \cdot \Phi(K-1+i-j)|$$

equals

$$\sum_{s=K-1}^{i-1} |A^{-1} e_K \cdot \Phi(s)|$$

The latter sum is bounded by

$$\sum_r |(A^{-1} e_K)_r| \sum_{s=K-1}^{i-1} |[\Phi(s)]_r|$$

where $(\quad)_r$ denotes the r th component of the vector.

We now notice that the condition that the roots be less than one in magnitude implies the convergence of

$$\sum_{s=K-1}^{\infty} |[\Phi(s)]_r|$$

for each r .

We now state the theorem which we have proved.

Theorem 4. Let y_i be the solution of

$$y_i = \sum_{j=1}^K a_j y_{i-j} + \sum_{j=0}^L b_j x_{i-j} + \delta_i$$

where the x_i 's are independent Gaussian random variables of mean zero and standard deviation σ . The δ_i 's are chosen as before. Moreover, assume that

$$\sum_j |a_j| < 1$$

Then there exists a constant N , independent of M and δ , such that

$$f(y_i - \tilde{y}_i)^2 d\mu \leq N \sup_j f\delta_j^2 d\mu$$

This last result shows that the solution to the discrete problem and to the continuous problem approach in the L_2 norm uniformly in the index as δ goes to zero and M goes to infinity. This is true because lemma 7 yields that

$$\sup_j (f\delta_j^2)$$

goes to zero as δ goes to zero and M goes to infinity. This is the main conclusion of this article. It allows a digital machine to be designed so that the output agrees with theory to any prescribed degree of accuracy. We would also, however, want the correlation function of the output to approximate that of the exactly calculated output. To this, we would have to bound

$$f\delta_i \delta_j d\mu$$

To get an estimate for $f\delta_i \delta_j d\mu$ which approaches zero as $i-j \rightarrow \infty$, analogous to lemmas 3 and 4, define

$$R_{k,m} = \{(x_{k+1}, \dots, x_{k+m+L}) : |\sum_{j=0}^L b_j x_{i+L-j}| < M(i+a), i = k+1, \dots, k+m\}$$

where m is any positive integer. The sets

$$R_{k+l(m+L),m}, \quad l = 0, 1, 2, \dots$$

are independent, and

$$P_m = \mu(R_{k,m})$$

is independent of k . Thus, an inequality similar to that of lemma 3 can be derived:

$$|\int \delta_j \delta_k d\mu| \leq \left\{ \frac{\nu}{(2\pi)^{1/2}} \exp \left[-\frac{M^2(1-a)^2}{2\nu^2} \right] + \delta \right\}^2 P_m^{\beta-1}$$

for $j > k + m + L$, where β is the integral part of $(j - k)/(m + L)$. Details are omitted.

1. The ϵ -Entropy of Certain Singular Measures on the Real Line, T. S. Pitcher⁴

1. Introduction

The ϵ -entropy of a probability distribution on a metric space was introduced in Ref. 1 for the purpose of defining data compression ratios. If C is a countable covering of the space by measurable sets, we write $\|C\| = \max_{A \in C} [\text{diameter}(A)]$, $\#(C)$ = number of sets in C and

$$H(C) = \sum_{A \in C} P(A) \log \frac{1}{P(A)}$$

Then the ϵ -entropy H_ϵ is given by

$$H_\epsilon = \inf_{\|C\| \leq \epsilon} H(C)$$

In this article we derive estimates of the asymptotic behavior of H_ϵ for certain singular measures on $[0, 1]$. The metric will be the usual length, and we will write $|A|$ for the length of an interval A . It is known (Ref. 1) that only coverings by intervals need be considered in computing H_ϵ .

It will be convenient to use the notation $\phi(x) = x \log 1/x$. The function ϕ is convex and has the property that: if

$$p_i \geq 0, \sum_{i=1}^n p_i = 1$$

then

$$\sum_{i=1}^n \phi(p_i) \leq \log n$$

The theorems of this article give asymptotic comparisons of H_ϵ with $\log 1/\epsilon$ which is the ϵ -entropy of Lebesgue measure on $[0, 1]$, plus a term which approaches 0 with ϵ .

⁴Consultant, Mathematics Department, University of Southern California, Los Angeles, Calif.

The asymptotic ratios are given in terms of various information-theoretic quantities.

2. Measures Related to N -adic Expansions

Let N be a fixed integer, $N \geq 2$, and let $(a_i, i = 1, 2, \dots)$ be a stationary stochastic process taking the values $0, 1, \dots, N-1$. We assume that no fixed sequence $a_i^0, i = 1, 2, \dots$ has positive probability. Define $k_i(x)$ for irrational x in $[0, 1]$ by

$$x = \sum_{i=1}^{\infty} k_i(x) N^{-i}$$

where the sum on the right is the N -adic expansion of x . Write

$$I_n(\ell_1, \dots, \ell_n) = [x | k_1(x) = \ell_1, \dots, k_n(x) = \ell_n]$$

and also

$$I_n(x) = I_n[k_1(x), \dots, k_n(x)]$$

The probability measure P associated with the process induces a measure, which we also call P , on $[0, 1]$ through the formula

$$P[I_n(\ell_1, \dots, \ell_n)] = P(a_1 = \ell_1, \dots, a_n = \ell_n)$$

According to the Shannon-McMillan-Breiman theorem (Ref. 2)

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P[I_n(x)] = -h(P) \text{ a.e. } (P)$$

where

$$h(P) = \sum_{i=0}^{N-1} P(a_1 = i) \log \frac{1}{P(a_1 = i)}$$

Let C_n be the covering of $[0, 1]$ by N -adic intervals, i.e., $C_n = \{J_k = [kN^{-n}, (k+1)N^{-n}], k = 0, 1, \dots, N^n - 1\}$.

Lemma 1. For $\epsilon = N^{-n}$,

$$\frac{1}{2} H(C_n) \leq H_\epsilon \leq H(C_n)$$

Proof. The second inequality is obvious from the definition of H_ϵ , since $\|C_n\| = N^{-n}$. Let $C = \{I_i\}$ be any

covering by intervals with $\|C\| \leq \epsilon$. If $I_{i_1}^k, \dots, I_{i_m}^k$ are the intervals intersecting J_k then

$$\sum_{j=1}^m \phi[P(I_{i_j}^k)] \geq \phi\left[\sum_{j=1}^m P(I_{i_j}^k)\right] \geq \phi[P(J_k)]$$

Summing on k and noting that each I_i hits at most three J_k 's, we have $2H(C) \geq H(C_n)$ from which the result follows. Lemma 1 is proved.

Now fix $\delta > 0$ and set

$$C'_n = \left[J_k \mid \left| \frac{1}{n} \log P(J_k) + h(P) \right| \leq \delta \right]$$

$$C''_n = \left[J_k \mid \left| \frac{1}{n} \log P(J_k) + h(P) \right| > \delta \right]$$

$$H'_n = H(C'_n)$$

$$H''_n = H(C''_n)$$

$$q(n, \delta) = P(\cup_{C'_n} J_k)$$

By Chung's theorem, stated above, $q(n, \delta)$ goes to 0 as n goes to ∞ .

Lemma 2.

$$H''_n \leq q(n, \delta) n \log N + \phi[q(n, \delta)]$$

Proof. Set $q = q(n, \delta)$ and

$$P_k = \begin{cases} P(J_k)/q, & \text{if } J_k \in C'_n \\ 0, & \text{if } J_k \in C''_n \end{cases}$$

Since $\sum P_k = 1$, we have

$$\begin{aligned} n \log N &\geq \sum_0^{N^n-1} \phi(P_k) = \frac{1}{q} \sum_{C'_n} \phi[P(J_k)] \\ &\quad + \frac{1}{q} \sum_{C''_n} P(J_k) \log q \\ &= \frac{1}{q} H''_n - \log \frac{1}{q} \end{aligned}$$

Theorem 1.

$$\frac{1}{2} \frac{h(P)}{\log N} \leq \lim_{\epsilon \rightarrow 0} \frac{H_\epsilon}{\log \frac{1}{\epsilon}} \leq \overline{\lim}_{\epsilon \rightarrow 0} \frac{H_\epsilon}{\log \frac{1}{\epsilon}} \leq \frac{h(P)}{\log N}$$

Proof. With the notation as above,

$$\begin{aligned} [1 - q(n, \delta)] n [h(P) - \delta] &\leq \sum_{C''_n} \phi[P(J_k)] \\ &= H'_n \\ &\leq [1 - q(n, \delta)] n [h(P) + \delta] \end{aligned}$$

Thus, since $q(n, \delta) \rightarrow 0$,

$$\lim_{n \rightarrow \infty} \frac{H(C_n)}{n \log N} \geq \lim_{n \rightarrow \infty} \frac{H'_n}{n \log N} \geq \frac{h(P) - \delta}{\log N}$$

and

$$\begin{aligned} \overline{\lim}_{n \rightarrow \infty} \frac{H(C_n)}{n \log N} &= \overline{\lim}_{n \rightarrow \infty} \frac{H'_n + H''_n}{n \log N} \\ &\leq \overline{\lim}_{n \rightarrow \infty} \frac{q(n, \delta) n \log N + \phi[q(n, \delta)] + [1 - q(n, \delta)] n (H\epsilon\delta)}{n \log N} \\ &= \frac{h(P) + \delta}{\log N} \end{aligned}$$

Since δ is arbitrarily small, this proves the theorem for the sequence $\epsilon_n = N^{-n}$.

In general, if we define $n(\epsilon)$ by

$$N^{-n(\epsilon)-1} \leq \epsilon < N^{-n(\epsilon)}$$

then

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \frac{H_{\epsilon}}{\log \frac{1}{\epsilon}} &\geq \lim_{\epsilon \rightarrow 0} \frac{H_N^{-n(\epsilon)}}{[n(\epsilon) + 1] \log N} \\ &= \lim_{n \rightarrow \infty} \frac{H_N^{-n}}{(n + 1) \log N} \geq \frac{1}{2} \frac{h(P)}{\log N} \end{aligned}$$

and

$$\lim_{\epsilon \rightarrow 0} \frac{H_{\epsilon}}{\log \frac{1}{\epsilon}} \leq \lim_{\epsilon \rightarrow 0} \frac{H_N^{-n(\epsilon)-1}}{n(\epsilon) \log N} = \lim_{n \rightarrow \infty} \frac{H_N^{-n-1}}{n \log N} \leq \frac{h(P)}{\log N}$$

Theorem 1 is proved.

3. Measures Related to Continued Fraction Expansions

Every irrational number in $[0, 1]$ has a unique infinite continued fraction expansion

$$x = \frac{1}{a_1(x) + \frac{1}{a_2(x) + \frac{1}{\ddots}}}$$

where the $a_i(x)$ are positive integers. We will write

$$I_n(k_1, \dots, k_n) = [x | a_1(x) = k_1, \dots, a_n(x) = k_n]$$

and

$$I_n(x) = I_n[a_1(x), \dots, a_n(x)]$$

If $(a_i, i = 1, 2, \dots)$ is a stationary ergodic process taking positive integer values then the probability measures P induces, as before, a measure P on $[0, 1]$ such that

$$P(I_n(k_1, \dots, k_n)) = P(a_1 = k_1, \dots, a_n = k_n)$$

As before, we assume that the induced measure has no atoms. We assume that

$$h(P) = \sum_{i=1}^{\infty} \phi[P(a_1 = i)] < \infty$$

$$h_0(P) = 2 \int_0^1 \log \frac{1}{t} P(dt) < \infty$$

Then by Chung's extension of the Shannon-McMillan-Breiman theorem (Ref. 2),

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(I_n(x)) = -h(P) \text{ a.e. } (P)$$

and by theorem 2.2 of Ref. 3 [with $f(x) = x^{-1}$]

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log |I_n(x)| = -h_0(P) \text{ a.e. } (P)$$

Fix $\delta > 0$ and set

$$C'_n = \left\{ I_n(k_1, \dots, k_n) \left| \frac{1}{n} \log P[I_n(k_1, \dots, k_n)] + h(P) \right| < \delta \right\}$$

and

$$\left\{ \frac{1}{n} \log |I_n(k_1, \dots, k_n)| + h_0(P) \right| < \delta \}$$

Then

$$\#(C'_n) \leq e^{n[h_0(P) + \delta]} + 1$$

and

$$\|C'_n\| \leq e^{-n[h_0(P) - \delta]}$$

We can find a covering C''_n of the remainder of $[0, 1]$ with

$$\#(C''_n) \leq 2(e^{n[h_0(P) + \delta]} + 1)$$

and

$$\|C''_n\| \leq e^{-n[h_0(P) - \delta]}$$

In fact, if we start with intervals of the form

$$[k e^{-n[h_0(P)-\delta]}, (k+1) e^{-n[h_0(P)-\delta]}]$$

and successively delete the intervals of C'_n , each deletion will add at most one interval and this will give the desired covering. Let C_n be the combined covering. For convenience, we take n so large that

$$2(e^{n[h_0(P)+\delta]} + 1) \leq e^{n[h_0(P)+2\delta]}$$

If we set $H'_n = H(C'_n)$ and $H''_n = H(C''_n)$, then we can prove, exactly as in lemma 2, that

$$H''_n \leq q(n, \delta) n [h_0(P) + 2\delta] + \phi[q(n, \delta)]$$

where

$$q(n, \delta) = \sum_{I \in C'_n} P(I)$$

goes to 0 as n goes to ∞ . We can now prove theorem 2.

Theorem 2.

$$\lim_{\epsilon \rightarrow 0} \frac{H_\epsilon}{\log \frac{1}{\epsilon}} \leq \frac{h(P)}{h_0(P)}$$

Proof. The proof is almost an exact duplicate of the corresponding part of the proof of theorem 1, and is omitted.

It is not possible to get the opposite inequality by the same device as before since an interval of length $e^{-n[h_0(P)-\delta]}$ could hit roughly $e^{2n\delta}$ intervals of C'_n .

4. The Case $h_0(P) = \infty$

In this part, we are concerned with the case where $(a_i, i = 1, \dots)$ is an integer-valued stationary ergodic process as in the previous part, with P the measure induced by the process through the continued fraction representation and where

$$h(P) = \sum_{i=1}^{\infty} \phi(P(a_1 = i)) < \infty$$

but

$$h_0(P) = 2 \int_0^1 \log \frac{1}{t} P(dt) = \infty$$

To see that such cases exist, note first that $h_0(P) = \infty$ if and only if

$$\int \log a_1(t) P(dt) = \infty$$

since

$$\log a_1(t) \leq \log \frac{1}{t} \leq \log [1 + a_1(t)] \leq 1 + \log a_1(t)$$

Thus, if we take the a_i to be independent with $P(a_1 = 1) = p_i$, we need only choose the p_i so that

$$\sum_{i=1}^{\infty} \phi(p_i) = h(P) < \infty$$

while

$$\sum_{i=1}^{\infty} p_i \log i = \int_0^1 \log a_1(t) P(dt) = \infty$$

We will need some facts about continued fraction expansion (Ref. 4). If we write

$$\frac{P_n(x)}{Q_n(x)} = \frac{1}{a_1(x) + \frac{1}{a_2(x) + \frac{1}{\ddots + \frac{1}{a_n(x)}}}}$$

then $P_n(x)$ and $Q_n(x)$ are generated by

$$P_0(x) = 0, Q_0(x) = 1$$

$$P_{n+1}(x) = a_{n+1}(x) P_n(x) + P_{n-1}(x)$$

$$Q_{n+1}(x) = a_{n+1}(x) Q_n(x) + Q_{n-1}(x)$$

and we have

$$|I_n(x)| = \frac{1}{Q_{n-1}(x) [Q_n(x) + Q_{n-1}(x)]}$$

Lemma 3. There exist numbers $A_N \uparrow \infty$ such that

$$\lim_{n \rightarrow \infty} \frac{1}{u} \log \frac{1}{|I_n(x)|} \geq A_N \text{ a.e. } (P)$$

Proof. Define A_N by

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{|I_n(x)|} &\geq \lim_{n \rightarrow \infty} \frac{2}{n} \log Q_{n-1}(x) \\ &\geq \lim_{n \rightarrow \infty} \frac{2}{n} \sum_{i=1}^{n-1} \log a_i(x) \end{aligned}$$

$$\begin{aligned} &\geq \lim_{n \rightarrow \infty} \frac{2}{n} \sum_{i=1}^{n-1} \log \{\min [a_i(x), N]\} \\ &= 2 \int_0^1 \log \{\min [a_1(x), N]\} P(dx) \end{aligned}$$

Now let, for fixed N and $\delta > 0$,

$$C'_n = \left\{ I_n(k_1, \dots, k_n) \mid \left| \frac{1}{n} \log P[I_n(k_1, \dots, k_n)] + h(P) \right| < \delta \right\}$$

and

$$\frac{1}{n} \log \frac{1}{|I_n(k_1, \dots, k_n)|} \geq A_N \}$$

$$q(n, \delta, N) = 1 - P(\cup_{C'_n} I_n)$$

By Chung's theorem and the lemma above, $q(n, \delta, N)$ goes to 0 as n goes to ∞ . Also,

$$\|C'_n\| \leq e^{-nA} N \quad \text{and} \quad \#(C'_n) \leq e^{n[h(P)+\delta]}$$

since

$$1 \geq \sum_{C'_n} P(I_n) \geq \#(C'_n) e^{-n[h(P)+\delta]}$$

As before, we can find a covering C''_n of the complement with

$$\|C''_n\| \leq e^{-nA} N$$

and

$$\#(C''_n) \leq e^{nA} N + 1 + e^{n[h(P)+\delta]}$$

We take C_n to be the combined covering and assume for convenience that N is so large that

$$e^{nA} N + 1 + e^{n[h(P)+\delta]} \leq e^{2nA} N$$

As before,

$$\begin{aligned} H''_n = H(C''_n) &\leq q(n, \delta, N) \log \#(C''_n) + \phi[q(n, \delta, N)] \\ &\leq q(n, \delta, N) 2nA_N + \phi[q(n, \delta, N)] \end{aligned}$$

and

$$H'_n = H(C'_n) \leq n[h(P) + \delta]$$

This gives, for $\epsilon = e^{-nA} N$,

$$\frac{H_{\epsilon}}{\log \frac{1}{\epsilon}} \leq \frac{H'_n + H''_n}{nA_N}$$

$$\leq \frac{h(P) + \delta}{A_N} + 2q(n, \delta, N) + \frac{\phi[q(n, \delta, N)]}{nA_N}$$

We can now proceed in the usual way to get the following extension of theorem 2.

Theorem 3. If the stationary ergodic process

$$(a_i, i = 1, 2, \dots)$$

has $h(P) < \infty$ but $h_0(P) = \infty$, then

$$H_{\epsilon} = o\left(\log \frac{1}{\epsilon}\right)$$

References

1. Posner, E., Rodemich, E., and Rumsey, H., "Epsilon Entropy of Stochastic Processes," *Ann. Math. Statist.*, Vol. 38, pp. 1000-1020, 1967.
2. Chung, K. L., "A Note on the Ergodic Theorem of Information Theory," *Ann. Math. Statist.*, Vol. 32, pp. 612-614, 1961.
3. Kinney, J., and Pitcher, T., "The Dimension of Some Sets Defined in Terms of f -expansions," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, Vol. 4, pp. 293-315, 1966.
4. Khintchine, A. Ya, *Continued Fractions*. P. Noordhoff, Ltd., The Netherlands.

XXI. Communications Elements Research

TELECOMMUNICATIONS DIVISION

A. RF Techniques: 90-GHz Millimeter Wave Work, W. V. T. Rusch,¹ S. D. Slobin,¹ and C. T. Stelzried

The objective of the millimeter wave work is to investigate millimeter wave components and techniques to ascertain the future applicability to space communications and tracking. This involves the development of instrumentation for accurate measurements of insertion loss, VSWR, power, and equivalent noise temperatures. Millimeter wave circuit elements are being evaluated in a radio telescope system consisting of a 60-in. antenna and a superheterodyne radiometer (SPS 37-33, Vol. IV, p. 245).

During the period of April 4 through July 10, 1966 a series of lunation observations was carried out to determine equivalent 90-GHz lunar disc brightness temperatures (SPS 37-45, Vol. IV, p. 313). On February 8, 1966, sixteen 3.3-mm observations of the sun were carried out to determine the equivalent blackbody disc temperature of the sun. This temperature was determined to be $6378.97 \pm 174.7^\circ\text{K}$ (pe). A series of measurements made on five subsequent days yielded an equivalent blackbody disc temperature of $6375.1 \pm 61.6^\circ\text{K}$ (pe).

¹Consultant from the University of Southern California, Electrical Engineering Department.

The technique used in observing the sun was identical to that described in the report of lunation observations (SPS 37-45, Vol. IV, p. 313). Eight observations prior to meridian transit on February 8, 1967 yielded $T'_s/T_{GT} = 3.07 \pm 0.16$ (pe), where T'_s is the equivalent blackbody antenna disc temperature of the sun, and T_{GT} is the equivalent excess noise temperature of the gas tube at the output of the waveguide switch. The eight observations following meridian transit yielded $T'_s/T_{GT} = 3.16 \pm 0.03$ (pe). Averaging these two values (to allow for the possibility of a uniform rate of change of atmospheric loss) yielded $T'_s/T_{GT} = 3.12 \pm 0.08$ (pe). Calibration of the equivalent excess noise temperature of the gas tube at the output of the waveguide switch yielded $T_{GT} = 1181.3 \pm 11.4^\circ\text{K}$. (The gas tube output passed through a 10-dB directional coupler into the main RF path.) This result then yielded $T'_s = 3683.2 \pm 100.9^\circ\text{K}$. The equivalent blackbody disc temperature of the sun T_s is then obtained by dividing T'_s by the BCF,² which was measured to be 0.58 for a solar radius of $16'15''$. The final result of this measurement was $T_s = 6378.97 \pm 174.7^\circ\text{K}$.

During the eight posttransit observations on February 8, the observations were also calibrated directly with the hot reference loads that were used to calibrate the gas tube, which served as a transfer standard. The results of this calibration technique yielded $T_s = 6277.6 \pm 230.7^\circ\text{K}$.

²Beam correction factor.

Following the observations of February 8, the Tucor, Inc. gas tube was replaced with an International Telephone and Telegraph Incorporated gas tube. It was expected that the new gas tube would provide more stability in the magnitude of the calibration pulses. Then on 5 days (February 12, 16, 17, 18, 19, 1967) pre- and posttransit observations were made. A typical set of data is plotted in Fig. 1. Averaging and reducing the data taken on these 5 days yielded:

February	T_s , °K	pe, °K
12	6372.7	89.5
16	6655.3	46.8
17	6272.0	41.9
18	6190.2	54.0
19	6331.2	90.6

The average of these five values, weighted inversely as the pe, is $T_s = 6375.1 \pm 61.6^\circ\text{K}$. This value is not far from the 3.2-mm value of 6402°K (Ref. 1). It should be noted that the pe quoted is statistical only. It does not include the large uncertainty involved in the determination of the BCF, which brings an additional uncertainty of 7-9%.

The findings mentioned in this article have been previously reported in a University of Southern California Electrical Engineering Report (Ref. 2). The results in the University of Southern California report have been referenced in an Aerospace Corporation report which surveys many solar disc temperature studies (Ref. 3). The value of $6375 \pm 61^\circ\text{K}$ compares favorably with values determined by other experimenters, as listed in Ref. 3.

References

1. Simon, M., *Solar Observations at 3.2 mm*, Technical Report CSUAC 10. Cornell-Sydney University Astronomy Center, Cornell University, Ithaca, N.Y., Nov. 1964.
2. Rusch, W. V. T., Slobin, S. D., Stelzried, C. T., *Millimeter Wave Radiometry for Radio Astronomy*, Final Report, USCEE 183. University of Southern California, Los Angeles, Calif., Dec. 1966.
3. Shimabukuro, F. I., and Stacey, J. M., *Brightness Temperature of the Quiet Sun at Centimeter and Millimeter Wavelengths ... and Chromospheric Models*, Technical Report TR-0158(3230-46)-1. Aerospace Corporation, Los Angeles, Calif., Oct. 1967.

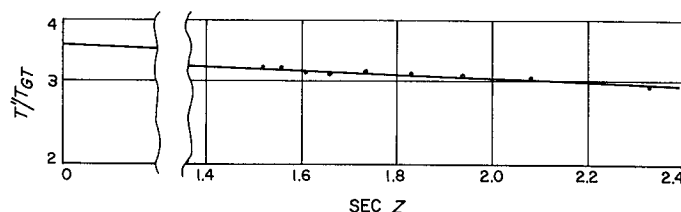


Fig. 1. February 16, 1966. Solar data (08:57-11:04 PST) with "best-fit" straight line approximation

$$T'_s/T_{6T} = 3.585 \pm 0.022 \text{ (pe);}$$

$$L_0 = 1.083 \pm 0.004 \text{ (pe)}$$

B. Quantum Electronics: Optical Communications Components, M. S. Shumate and J. C. Siddoway

1. Carbon Dioxide Laser Heterodyne Receiver, M. S. Shumate

The construction of two carbon dioxide lasers (SPS 37-45, Vol. IV, p. 316) has been completed, and extensive frequency drift tests have been performed. First attempts to stabilize the carbon dioxide laser to the line center have also been performed. Accurate measurements of the carbon dioxide laser wavelengths have been postponed temporarily by lead screw errors in the SPEX monochromator being used for the purpose.

a. Laser construction. The proper design of the laser cavity leads to its ultimate stability (Ref. 1). Ideally, the optical path length between the laser mirrors should not change at all. In practice, temperature changes lead to expansion or contraction of the mirror spacer; mechanical vibrations lead to flexing of the mirror spacer and relative tilting of the mirrors; and air currents through the cavity lead to small path length changes due to small fluctuations in the air temperature. The present laser design being used for the heterodyne receiver is a compromise in order to keep costs low and to reduce the fabrication time. The mirror mounts are commercially available, and the mirror spacer is made of Pyrex. The laser discharge tube is the Brewster Window type with extension tubes to cut down air currents between the windows and the mirrors. The output mirror is a flat with a dielectric film reflective coating that transmits 2%, and the other mirror is spherical with a coating that reflects $\sim 99\%$ at $10 \mu\text{m}$ and transmits partially in the visible.

Frequency drift tests on these lasers have been performed, and reveal that, after a 1-h warm-up, the drift is low enough to permit extended operation on one oscillating line without retuning. Precise drift tests cannot be performed until heterodyne operation is attained,

which has been delayed by the necessary repair of our high speed copper-doped germanium detector.

b. Laser frequency stabilization. A laser frequency stabilization technique (Ref. 2), which will stabilize to the center of the doppler broadened line which is lasing, has been successfully tested. Figure 2 presents a diagram of the apparatus. The lock-in amplifier, PAR Model HR-8, is basically a tuned amplifier followed by a demodulator/filter. The amplifier supplies a reference frequency which is added to the error signal and causes the piezoelectric translator to dither the mirror back and forth slightly, which changes the laser frequency. Since the gain of the laser changes due to the line shape of the transition which is lasing, the power out of the laser will be a maximum at the line center and will decrease on either side of the line center. By proper adjustment of the phase controls of the lock-in amplifier, the error signal can be made proportional to the slope of the gain curve of the laser. Thus the error signal is proportional to the offset of the average laser frequency from the frequency of the line center; and, when the loop is closed, the average laser frequency (or, more specifically, the carrier frequency of the FM signal) is then locked to the line center frequency. The present apparatus is being refined and a more sensitive detector will be incorporated.

The major source of difficulty with this technique, as applied to carbon dioxide lasers, is the multitude of different lines which are capable of oscillation. Lack of sufficient isolation between adjacent lines produces a tuning characteristic that may not have a peak at the line center, hence producing an undesirable offset during closed loop operation. This problem can be corrected when the accurate wavelength determinations have been made, thus giving an optimum cavity length.

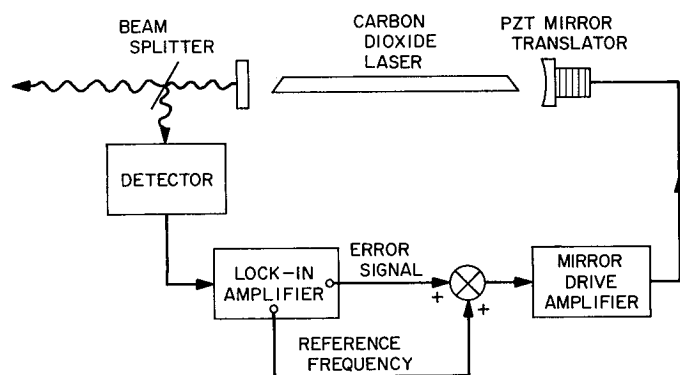


Fig. 2. Carbon dioxide laser frequency stabilization apparatus

c. Optimum cavity length determination. The optimum cavity length determination (SPS 37-39, Vol. IV, p. 196) has been delayed due to lack of high accuracy carbon dioxide laser wavelength information. An attempt has been made to perform the wavelength measurements using our SPEX grating monochromator. The method is as follows: very accurate reference points can be obtained by using high order lines from the helium-neon laser; measurements in the regions between the helium-neon reference points depend upon the accuracy of the lead screw. There is apparently a short term lead screw error which prevents interpolation to the desired accuracy.

2. Infrared Atmospheric Propagation Study, M. S. Shumate

An attempt is being made by J. A. Westphal of the California Institute of Technology Division of Geological Sciences to determine the effects of atmospheric turbulence on observations of stellar infrared sources. Since the results of this study will be very useful for understanding the effects of the atmosphere on a carbon dioxide laser beam, JPL is providing financial and technical support for this effort. Preliminary measurements have been performed, and an example is presented below.

The objective is to determine the apparent diameter of the blur circle at the Cassegrain focus of the 200-in. Hale telescope at several wavelengths in the infrared portion of the spectrum. The approach is to scan the lighted limb of the first quarter moon with an infrared photometer (Ref. 3) and record the outputs from the infrared detector and a visible detector. The infrared detector is equipped with a filter wheel containing several optical filters for the following wavelengths: 1.65, 2.2, 3.5, 5, 9, 11, and 13 μm ; there is also a broadband filter for the 8–14 μm range. The two detectors operate through two small adjacent focal plane apertures to limit the fields of view. Scanning is accomplished by driving the telescope at sidereal rate and allowing the moon to move out of the field of view ($\sim 1/2''/\text{s}$). The outputs of the two detectors are recorded on an FM instrumentation recorder, and are then sampled and digitized in the JPL Data Analysis Laboratory.

A series of runs has been made, with one of the typical scans shown in Fig. 3. The time shift between the two sets of data is caused by a misalignment of the pair of apertures. Interpretation of the fluctuations on both sets of the data is difficult because of the aperture alignment problem. A new photometer, with one common aperture for both detectors, is being fabricated at the present time. It will be used in all future measurements, and should result in higher resolution of the infrared images.

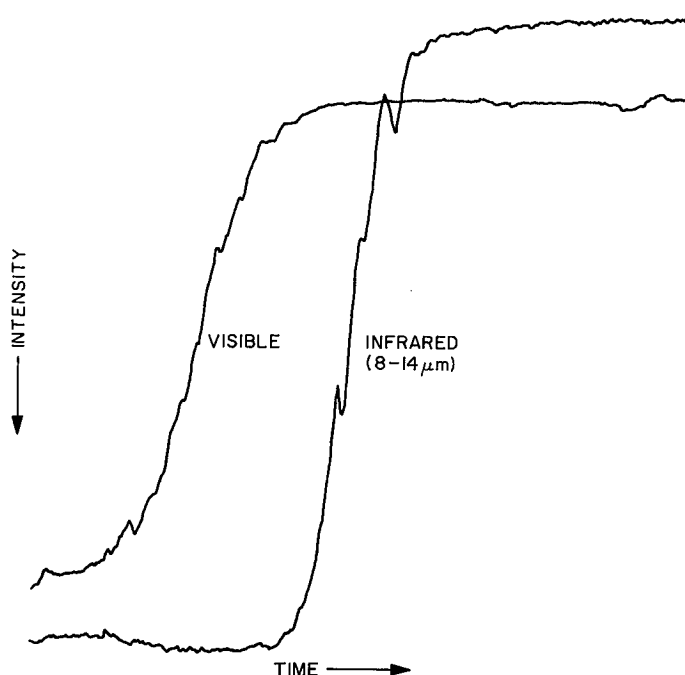


Fig. 3. Typical moon limb scan

3. Isotopic CO₂ Laser Studies, J. C. Siddoway

a. Introduction. This article is on the continuation of the work discussed in SPS 37-45, Vol. IV, p. 317, which described the construction and purpose of a CO₂ isotopic laser. The apparatus is being used in the studies of:

- (1) Impurity and/or "gettering" effects which limit the lifetime of sealed off CO₂ lasers.
- (2) Isotopic wavelength shifts to prevent radiative absorption by thermally excited CO₂ in the atmosphere.

Wavelength shifts have been measured in two of the three isotopes being used. A mass spectrometer analysis of the gas products from the laser is being used in lifetime studies. Presently, results indicate no gross impurity levels or unusual molecular species to limit lifetimes. Indications are it is primarily a slow depletion of CO₂ through sputtering and dissociation.

b. Experimental apparatus and procedures. To review briefly: the apparatus consists of a gas handling system and discharge tube, two high vacuum pumping systems, and a quadrupole mass spectrometer. High vacuum fittings are used throughout the system. The electrodes are internal hollow cylinders along the axis of the discharge tube and can be changed to evaluate different materials. The discharge tube has potassium chloride Brewster

windows sealed with low vapor pressure epoxy. Gold coated mirrors are used in a hemispherical configuration, with output coupling provided by a hole in the flat mirror.

An operational "run" consists of the following steps:

- (1) Evacuating the system to 10⁻⁸ torr or less and recording the background spectrum with the mass spectrometer.
- (2) Gases are then admitted into the discharge tube and at least two spectra recorded at different gain levels. This allows comparison of low level impurity peaks as well as the main constituents.
- (3) Exciting the discharge tube and recording the mass spectra to observe the immediate change in the initial gas constituents.
- (4) Wavelength measurements with the SPEX 0.75-m Czerny-Turner monochromator.
- (5) Continued operation, with periodic mass spectra scans, until laser action ceases.

In this manner the relative heights of the mass peaks can be compared throughout the lifetime of the run. Ambiguous molecules with the same mass (e.g. C¹²O¹⁶, N₂¹⁴; mass 28) can be resolved with the isotopic CO₂.

c. Experimental results—lifetime studies. Three isotopes of CO₂ (C¹²O₂¹⁸, C¹³O₂¹⁶, C¹⁴O₂¹⁶) are currently being used. Figure 4 shows the mass spectra obtained with helium and C¹³O₂¹⁶ at two different gain settings. From left to right the spectra represent background, gas composition before the discharge was run, and composition after laser action ceased. The outstanding features of the spectra are the decrease of the 44-45 peak (CO₂) and the increase in 28-29 peak (CO). This is typical of the other isotopes as well as ordinary CO₂.

Comparisons of all the mass spectra recordings do not reveal formation of any unusual molecular species or impurity concentration. The lifetime limitation simply appears to be due to dissociation of the CO₂ into CO and free oxygen, and cleanup of the oxygen by absorption, sputtering and chemical reaction. The pressure in the discharge tube is also observed to slowly decrease and is reduced as much as 15-20% when laser action stops. Addition of CO₂ to the discharge then restores laser action, and the process is repeated.

Nonreactive electrode materials, e.g., platinum, have increased the operational lifetime, but sputtering is still

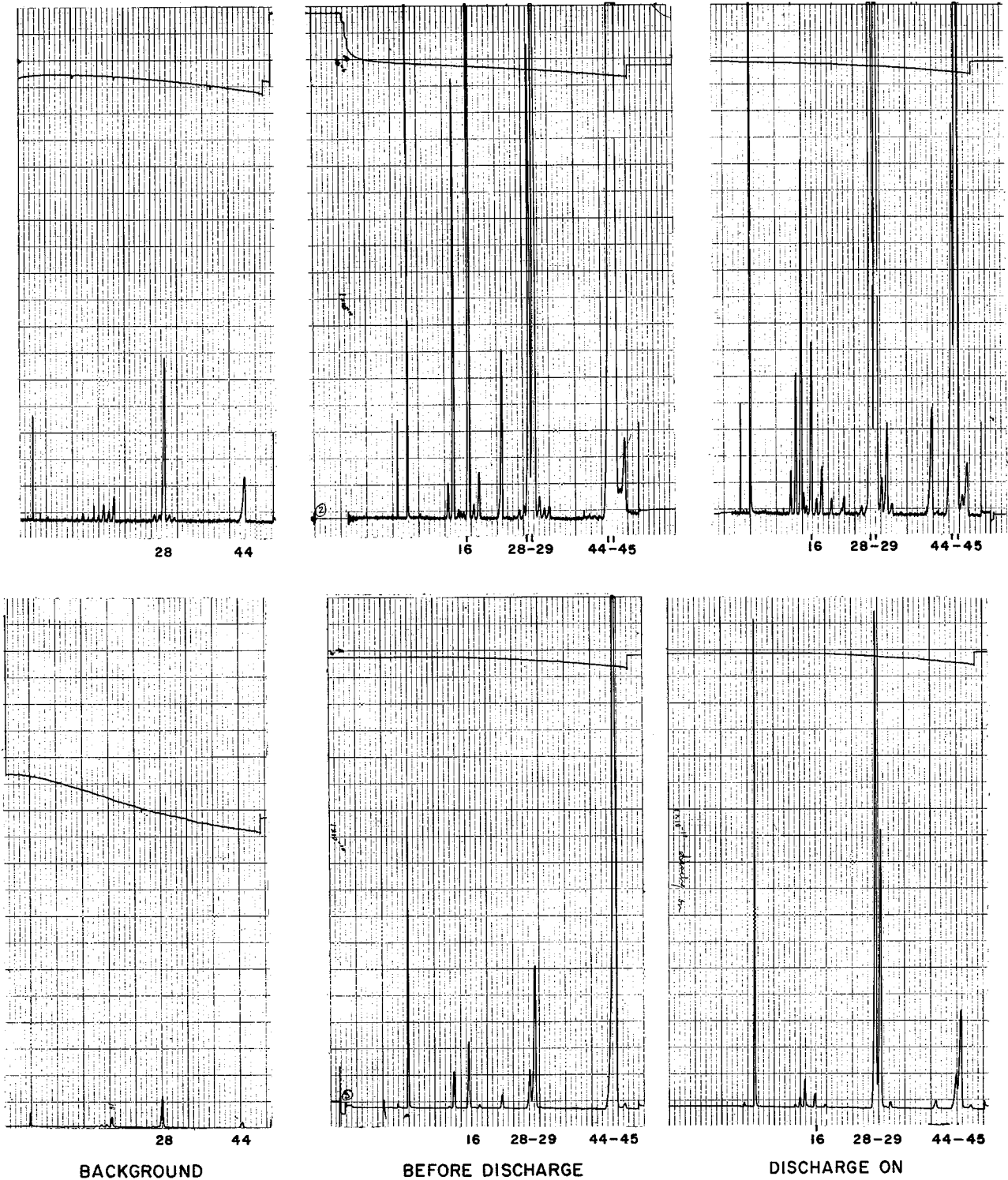


Fig. 4. Samples of $\text{He-C}^{13}\text{O}_2^{16}$ mass spectra

a major problem. Clark and Wada³ have reported that a new cathode design and addition of ~ 1 torr of xenon produced remarkable improvements in lifetime performance (2800 h). Modification of the electrode design is now being planned for the isotopic laser.

d. Wavelength measurements. Tables 1 and 2 show the laser transitions that have been measured for $C^{12}O_2^{18}$ and $C^{13}O_2^{16}$ respectively. Five lines of the O^{18} isotope were previously reported in the literature (Ref. 4). These were used for a calibration check and as a starting point to search for other lines. Two additional transitions have been measured. Table 2 lists 13 lines that have been observed with $C^{13}O_2^{16}$. These are compared with calculated values taken from Ref. 5.

No results have been observed with the C^{14} isotope. The specific activity, or purity, of the material is the

³P. O. Clark and J. Y. Wada, "Characteristics of CO_2 -Xe-He Lasers," to be published.

Table 1. Comparison of measured $C^{12}O_2^{18}$ transitions ($00^\circ 1-10^\circ 0$)

Identification	$\nu, \text{cm}^{-1}, \text{vac}$	
	Ref. 4	Measured
P(16)	—	1072.1
P(18)	1070.6	1070.3
P(20)	1069.0	1068.9
P(22)	1067.4	1067.4
P(24)	1065.8	1065.8
P(26)	1064.2	1064.2
P(28)	—	—
P(30)	—	1060.1

Table 2. Calculated and measured $C^{13}O_2^{16}$ transitions, cm^{-1} ($00^\circ 1-10^\circ 0$)

J	Calculated (Ref. 5)		Measured	
	P branch	R branch	P branch	R branch
12	903.5	—	903.60	—
14	901.9	924.3	901.87	924.32
16	900.2	925.7	900.18	925.72
18	898.5	927.1	898.43	927.13
20	896.7	928.4	896.69	928.44
22	895.0	—	894.93	—
24	893.2	—	893.17	—
26	891.4	—	891.35	—
28	889.6	—	889.54	—

highest available; however, it still contains appreciable amounts of $C^{12}O_2^{16}$ ($\sim 35\%$). Further work is planned using an optical resonator with less output coupling ($\sim 1\%$) to extend the measurements over weaker transitions.

References

1. *Quiet Laser Program*, Engineering Report 8639. Perkin Elmer Corporation, Norwalk, Conn., Mar. 16, 1967.
2. White, A. D., "Frequency Stabilization of Gas Lasers," *IEEE Trans. Quantum Electron*, QE-1, p. 349, Nov. 1965.
3. Westphal, J. A., Murray, B. C., and Martz, D. E., "An 8-14 Micron Infrared Astronomical Photometer," *Applied Optics*, Vol. 2, p. 749, July 1963.
4. Wieder, I., and McCurdy, G. B., "Isotope Shifts and the Role of Fermi Resonance in the CO_2 Infrared Maser," *Phys. Rev. Lett.*, Vol. 16, No. 13, Mar. 28, 1966.
5. Jacobs, G. B., and Bowers, H. C., "Extension of CO_2 -Laser Wavelength Range with Isotopes," *J. Appl. Phys.*, Vol. 38, pp. 2692-2693, May 1967.

C. Low Noise Transponder Preamplifier Research, S. M. Petty

The parametric amplifier has been considered for possible application as a low noise spacecraft transponder preamplifier. This type of amplifier can provide substantially lower noise figures at S-band than either transistor or tunnel diode amplifiers, provided that the restrictions imposed by a spacecraft environment can be met. These restrictions create two major design problems for a parametric amplifier:

- (1) The four-port circulator must be a miniature, light-weight design. Lately, circulators of this type have become available with the advent of smaller magnets and new stripline techniques.
- (2) The pump source must have a higher reliability under severe vibration as well as higher dc-to-RF efficiency than has been obtainable from klystron oscillators. Thus some type of solid state source is required.

A new parametric amplifier (American Electronics Laboratories model PAR 1612A) has been purchased with a solid state avalanche diode oscillator (Refs. 1, 2) as a pump source. This oscillator operates at approximately 13 GHz with a dc-to-RF efficiency of 5%. The size and weight of the oscillator along with its need for only one low voltage power supply makes it ideal for spacecraft application if the reliability could be proven satisfactory.

Ambient temperature performance of the amplifier is shown in Table 3. Figure 5 shows the entire parametric amplifier package including the miniature four-port circulator.

During laboratory testing, the avalanche diode oscillator failed after 25 h of operation at ambient temperature. Similar avalanche diode oscillators have been found to have an unpredictable life span before burnout. Since no method has yet been devised to test these diodes in a nondestructive manner, the avalanche diode oscillator is impractical for serious applications at the present time. Until a more reliable solid state pump source is found, the parametric amplifier cannot be considered feasible for use as a spacecraft transponder preamplifier.

Table 3. Parametric amplifier performance (American Electronics Laboratories model PAR 1612A)

Signal frequency	2115 MHz
Net gain	17 dB
Noise figure ^a	2.85 dB
Bandwidth (3 dB)	30 MHz
Temperature	25° C
Direct current voltage required	51 V
Direct current required	20 mA
Signal power for -1 dB gain compression	-38 dBm

^aMeasured with ambient and cooled terminations. Automatic noise figure meter measures 2.5 dB.

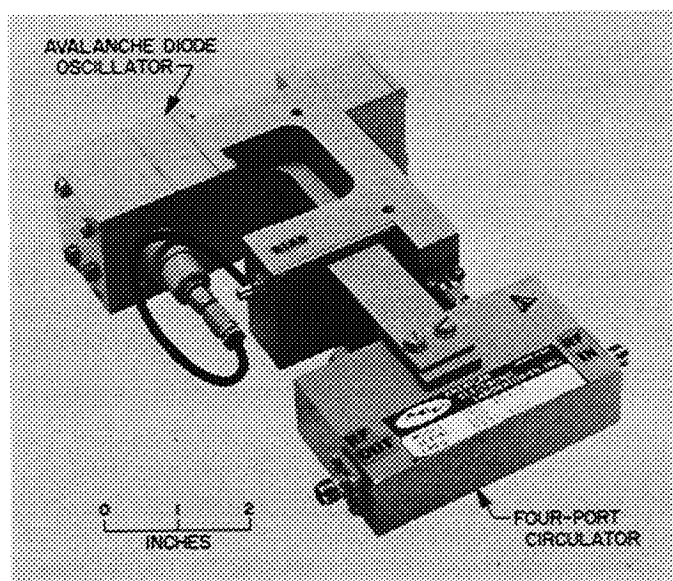


Fig. 5. AEL parametric amplifier with solid state pump source

Work is continuing on the tunnel diode amplifier described earlier (SPS 37-46, Vol. IV, pp. 251-252). This seems to be the most promising approach to this project.

References

1. Smith, K. D., "Generating Power at Gigahertz with Avalanche-Transit Time Diodes," *Electronics*, Vol. 39, No. 16, Aug. 8, 1966, pp. 126-131.
2. DeLoach, B. C., and Johnston, R. L., "Avalanche Transit-Time Microwave Oscillators and Amplifiers," *IEEE Trans. Electron Devices*, Vol. ED-13, No. 1, Jan. 1966, pp. 181-186.

D. Spacecraft Antenna Research, R. M. Dickinson and K. Woo

1. Antenna Pattern Tolerances, R. M. Dickinson

a. Introduction. The object of this study is to increase the accuracy of full scale spacecraft antenna pattern measurements. Antenna patterns recorded from full scale spacecraft antenna models are used in communications analysis and prediction.⁴

This article will present the overall tolerances to date, the individual tolerance contributors, and the current investigation to reduce tolerances.

b. Overall tolerances. The overall antenna pattern tolerances that have been achieved for spacecraft and the DSIF are shown in Fig. 6, as a function of the nominal pattern gain level. It can be seen that the symmetrical tolerance magnitude is generally an inverse function of the gain level. This result is principally due to the difficulty of creating a plane wave incident upon the spacecraft model in the presence of the earth and the model support and positioning structure.

Also, the resulting tolerances are secondarily influenced by a decreasing signal-to-noise ratio in the measuring and recording instrumentation at the lower gain levels. The constant portion of the tolerances stems from the inherent uncertainties in any measurement, such as absolute gain determination and the necessary system considerations. That is, any one individual antenna pattern can be more accurately determined than the results shown in Fig. 6; however, when all the actual operating conditions are included, the final tolerances must be increased to cover these conditions.

⁴J. S. Omahen and E. F. Oliver, "Telecommunications Prediction Program (EDCOM), PD900-47, Jet Propulsion Laboratory, Pasadena, Calif. Aug. 15, 1967.

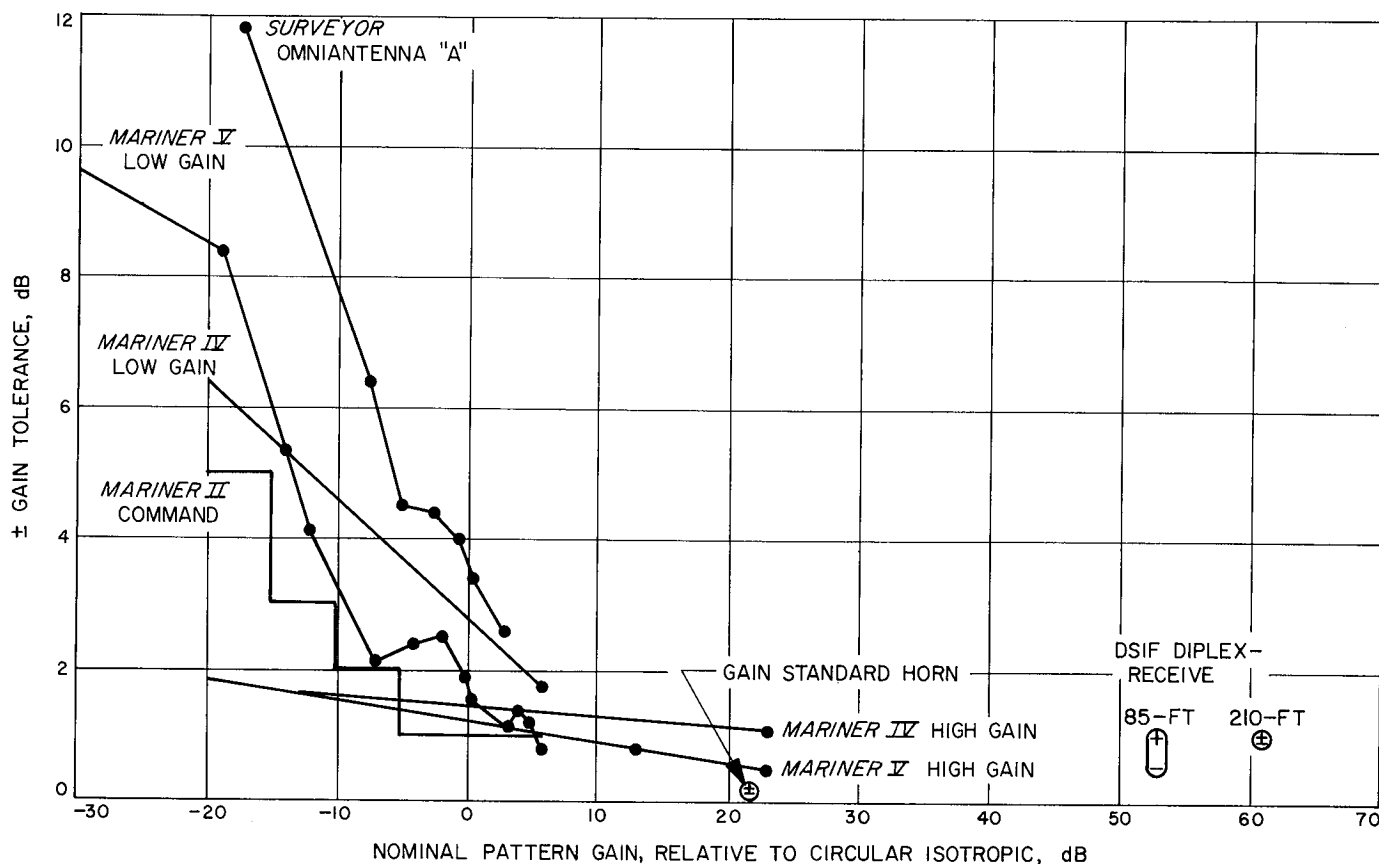


Fig. 6. Spacecraft antenna pattern tolerances

For example, measurements of the actual *flight* antennas on the actual *flight* spacecraft before flight are not possible. Instead, measurements are made of a *model* antenna on a *model* spacecraft in the 1-g field near the earth.

The *Mariner II* command antenna tolerances are lowest because they were simply estimated. The other spacecraft antenna tolerances are larger because they result from a linear summation of many individual measured, as well as estimated, contributors. The *Surveyor* tolerances were largest because of extremely low gain antennas.

c. Tolerance contributors. Table 4 is an example of the tolerance contributors considered for the *Mariner IV* low gain antenna. The dB/dB units refer to dB per dB down from the peak gain of the pattern. Similar tolerance contributors, suitably modified for particular mission system considerations, were obtained for the other spacecraft. Items 1 through 8 are generally self explanatory. Item 9 is a measure of the degree to which a plane wave is incident upon the model. Item 9 was obtained by field probing height-gain measurements and subtracting the difference between the same antenna's pattern recorded at different

positions relative to the range (on a perfect range, the patterns should be identical).

Item 10 is a measure of the accuracy to which absolute gain relative to circular isotropic can be established on a gain standard antenna.

To obtain the gain of the model antennas on the model spacecraft, the gain standard in item 10 is taken from its relatively short, clean range to the required long, corrupted spacecraft range where its output is padded down to a level approximately equal to the spacecraft antenna. The difference is then measured. Item 11 is the resulting accuracy of this measurement.

Items 12 and 13 are used to account for the fact that the possible flight antennas have different gains and pattern shapes, due to manufacturing tolerances.

Items 14 and 15 are used to account for the difference between the model on the range and the flight spacecraft. An indication of these differences was obtained by measuring the difference in gain of the same antenna on a very

Table 4. Mariner IV low gain antenna pattern tolerance contributors

Source ^a	Item	Value
m	1. Recording system linearity	± 0.015 dB/dB from peak
m	2. Illuminator and recording system stability	± 0.07 dB
m	3. Illuminator ellipticity stability	Negligible
m	4. Model positioner slip ring noise	± 0.01 dB/dB
m	5. Model rotary joint wow	± 0.10 dB
e	6. Illuminator-test antenna interaction	Negligible
e	7. Thermal and 1-g distortion of antenna	Negligible
m	8. Wind modulation of illuminator and model	10 mph, negligible
m	9. Range reflection and diffraction	$\pm (0.5 \text{ dB} + 0.128 \text{ dB/dB})$
m-e	10. Gain standard absolute calibration	± 0.10 dB
m-e	11. Comparison gain measurement	± 0.20 dB
m	12. Antenna peak gain differences	± 0.11 dB
m	13. Antenna pattern shape differences	± 0.025 dB/dB
m-e	14. Model-spacecraft differences	± 0.5 dB
e	15. Spacecraft-spacecraft differences	± 0.2 dB
m	16. Solar panel position (0, ± 0.20 – 1.20 deg)	Negligible
m	17. Solar panel damper interaction (43, ± 0 – 90 deg)	Negligible
m	18. Solar sail position (35 ± 20 deg)	Negligible
	Linear sum total	$\pm (1.78 + 0.178 \text{ dB/dB})$

^am = measured e = estimated

crude model and the better quality, but not flight quality, full scale spacecraft antenna pattern model.

d. Tolerance reduction. The current investigation to reduce tolerances began with instrumenting the range cherry-picker to perform high gain, height-gain field probing in order to reduce item 9 in Table 4 as applied to the spacecraft model high gain antenna. Figure 7 shows the range cherry picker instrumented with probing antennas, a distance measuring device and a recorder. After using the equipment to align the remote illuminator on the spacecraft location, the field strength was mapped horizontally and vertically at various look angles around the model location. The results are that with respect to a high gain antenna of the *Mariner* Mars 1969 class, the illuminator field strength is uniform to within less than ± 0.25 dB (recorder resolution).

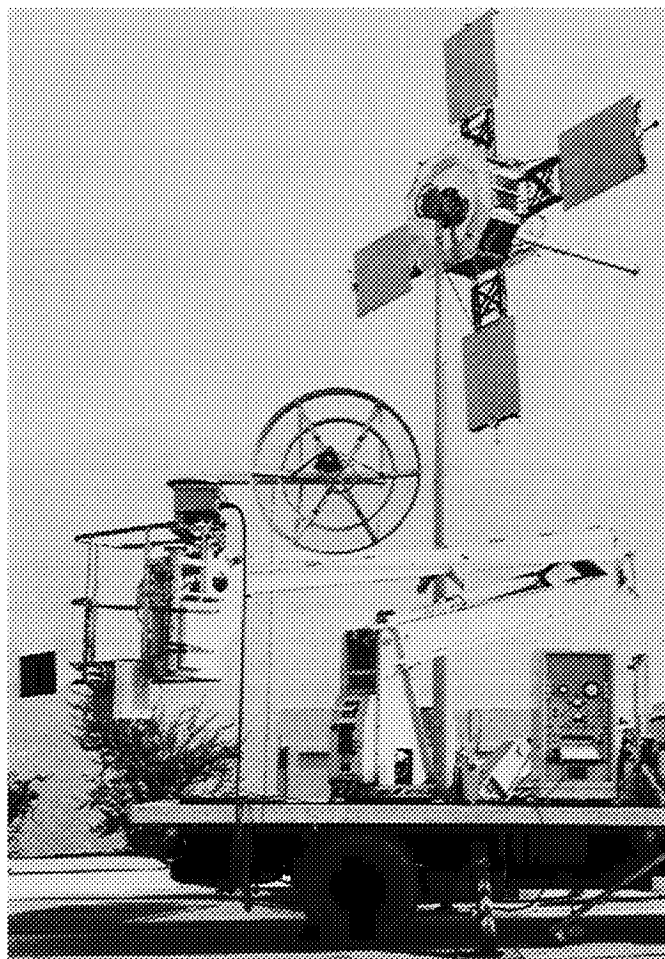


Fig. 7. High gain field probing instrumented cherry picker

Item 5 in Table 4 was attacked next. The equipment set-up of Fig. 8 was used to provide the high resolution required to measure the small rotary joint wow. The results of testing approximately 20 rotary joints, both new and used of three manufacturers, are that selected rotary joints (five) of a particular manufacturer have wow or transmission amplitude variations of less than ± 0.02 dB.

In order to extend the dynamic range of pattern recording and to provide for a higher modulation frequency for increased recording data rate, an investigation of the infinite impedance detector (Ref. 1) is underway. Figure 9 shows a preliminary characteristic detection curve as compared with the bolometer detectors now in use. Figure 10 shows the pencil triode tube outside the resonant cavity housing. If the detector can be made stable and repeatable the non-linearities can be corrected in the computer program that processes the recorded pattern tapes.

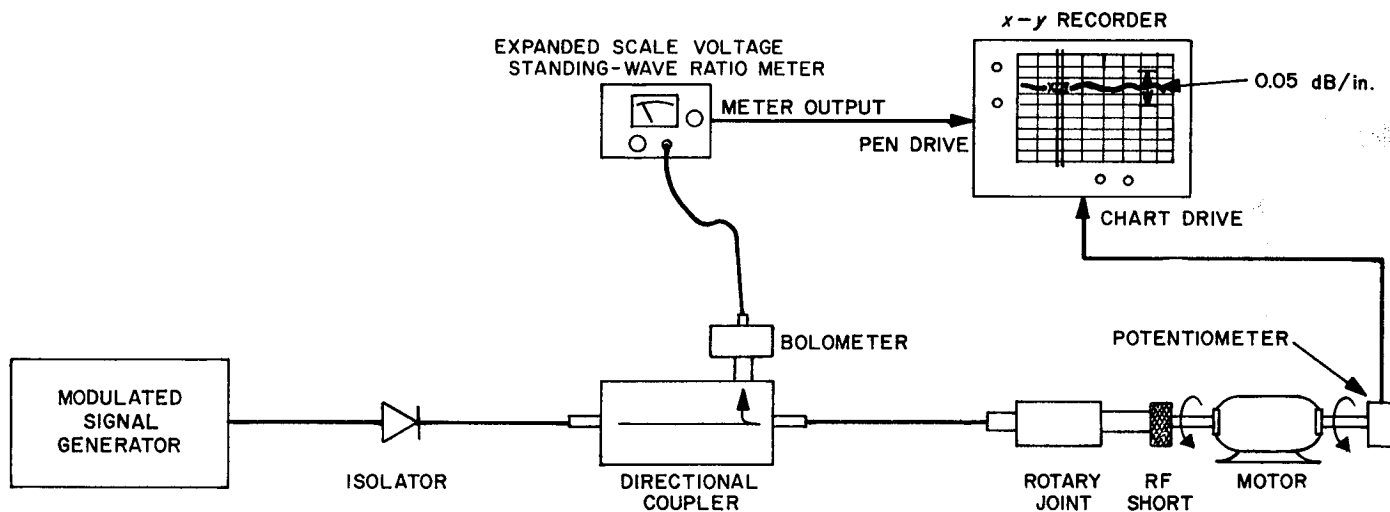


Fig. 8. Rotary joint wow test set-up

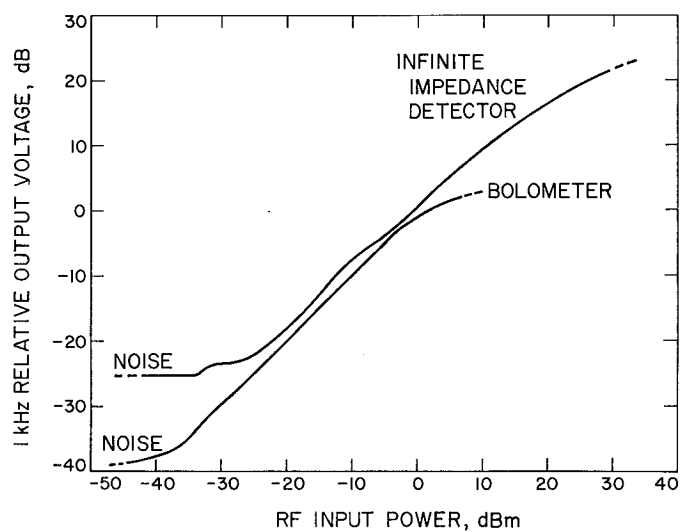


Fig. 9. Infinite impedance detector characteristic curve

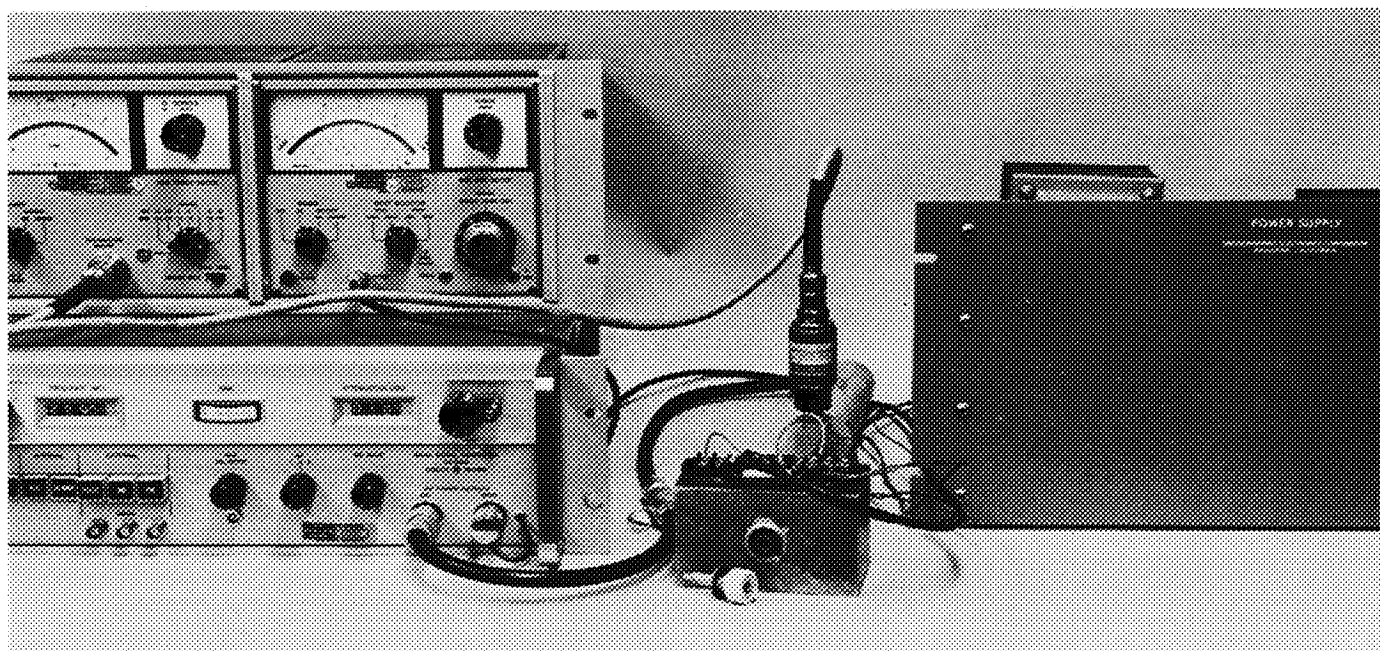


Fig. 10. Infinite impedance detector tube and cavity

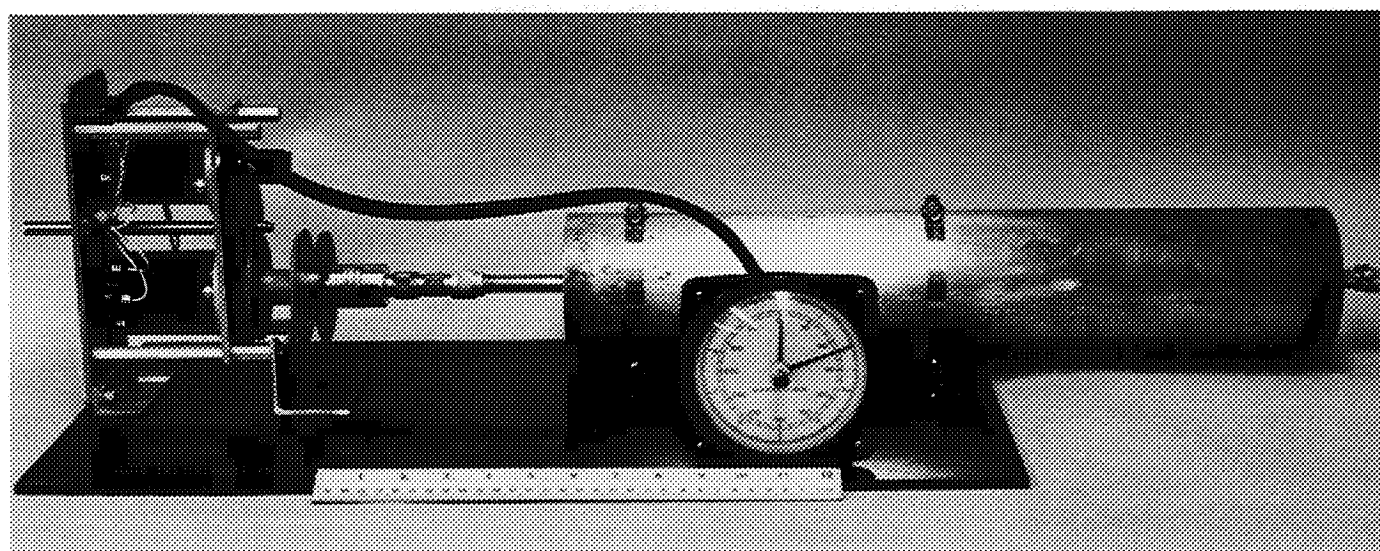


Fig. 11. The rotary antenna attenuator

The pattern tape recording program will also be modified to allow the recording of a reference amplitude signal at the end of each pattern cut in order to monitor the recording system gain drift (Item 2 Table 4) which can be subtracted out in the computer processing.

A compact, more accurate RF insertion loss creating device is needed for use in absolute gain calibrations, comparison gain measurements, and recording system linearity checks. At present, step attenuators are being

used. The existing models of the very accurate S-band rotary vane attenuator are rather cumbersome for range use. A preliminary model of a rotary antenna attenuator has been constructed. This device, shown in Fig. 11, creates RF attenuation by polarization mismatch loss between linearly polarized antennas mounted in circular waveguide. A synchro transmitter and indicator are used to read rotation angle. One antenna is rotatable, and a second fixed antenna supports a third terminated orthogonal antenna to absorb the cross polarized signal.

Figure 12 compares the measured attenuation with the theoretical loss curve in dB of $20 \log_{10} \sec \theta$, where θ is the rotation angle of the rotary antenna relative to the fixed antenna. The reflected power at the input was below -31.5 dB throughout the 90-deg rotation.

2. 400-MHz Coaxial Cavity Radiator, K. Woo

a. Introduction. This article reports the preliminary results of a 400-MHz, low gain, circularly polarized antenna designed for capsule relay-link communications.

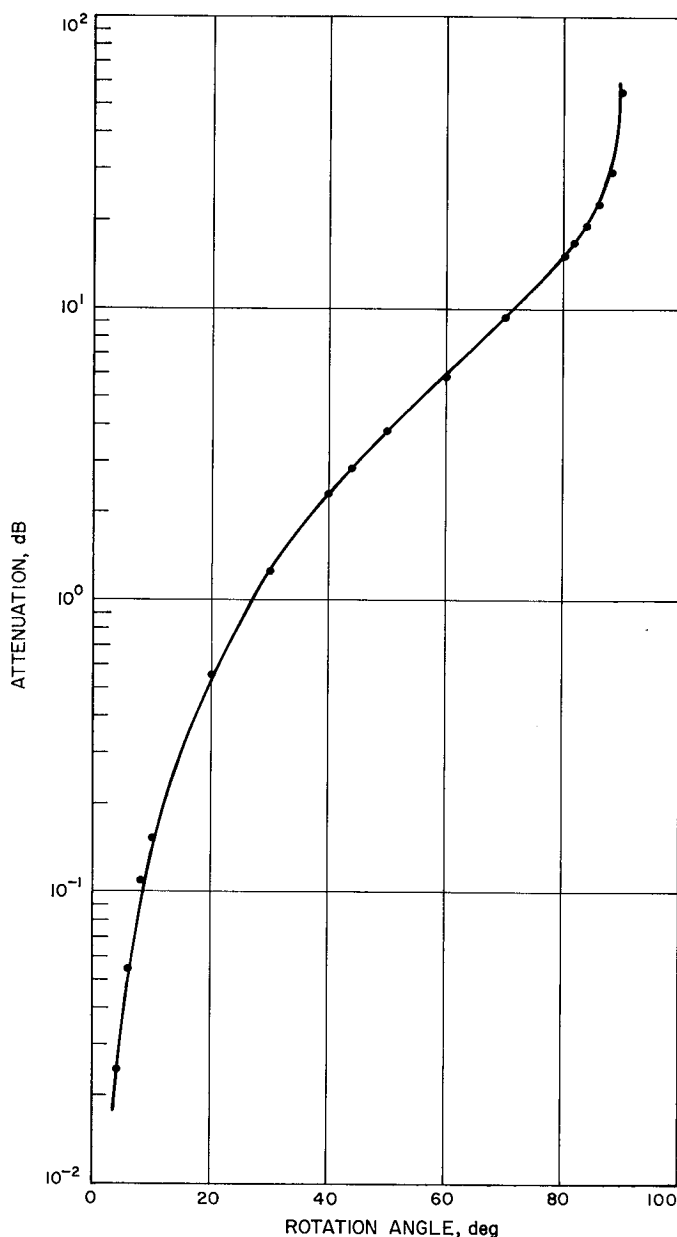


Fig. 12. Measured attenuation vs theoretical attenuation for the rotary antenna attenuator

The antenna is a coaxial cavity radiator (SPS 37-40, Vol. IV, pp. 201-206) operating on the TE_{11}^0 mode. It has a symmetrical beam shape and is sterilizable.

b. Antenna design. The experimental model of the antenna is shown in Fig. 13. It is composed of a coaxial cavity, open at one end and shorted at the other. The dimensions of the cavity, as shown in Fig. 14, have been properly selected that only the TE_{11}^0 mode will resonate at 400 MHz. The cavity is fed by two spatially orthogonal probes near the shorted end. The probes are each connected to an output terminal of a 3-dB hybrid (not shown). The function of the hybrid is to divide incoming energy from the feed line such that each probe receives an equal amount of power in time quadrature. When excited, the

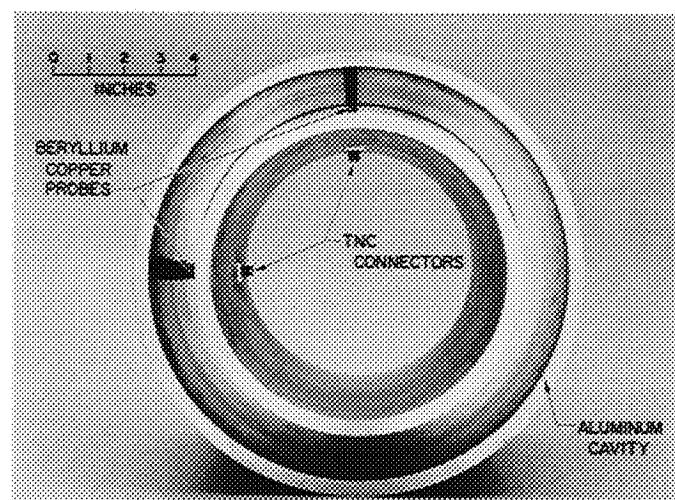


Fig. 13. Experimental model

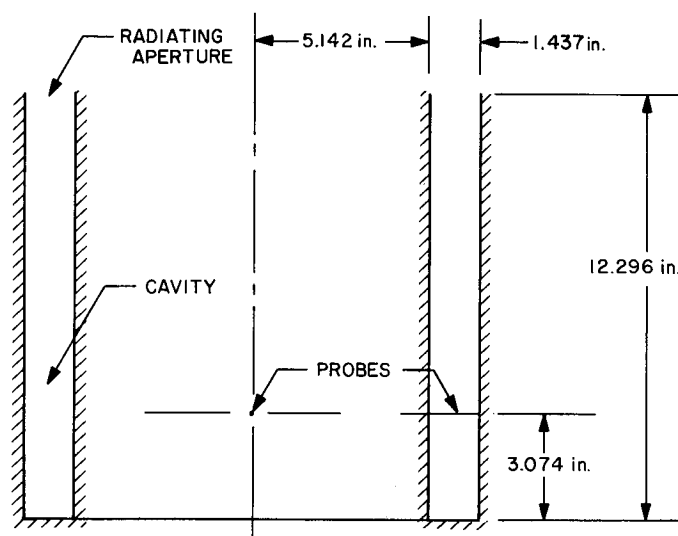


Fig. 14. Cavity dimensions

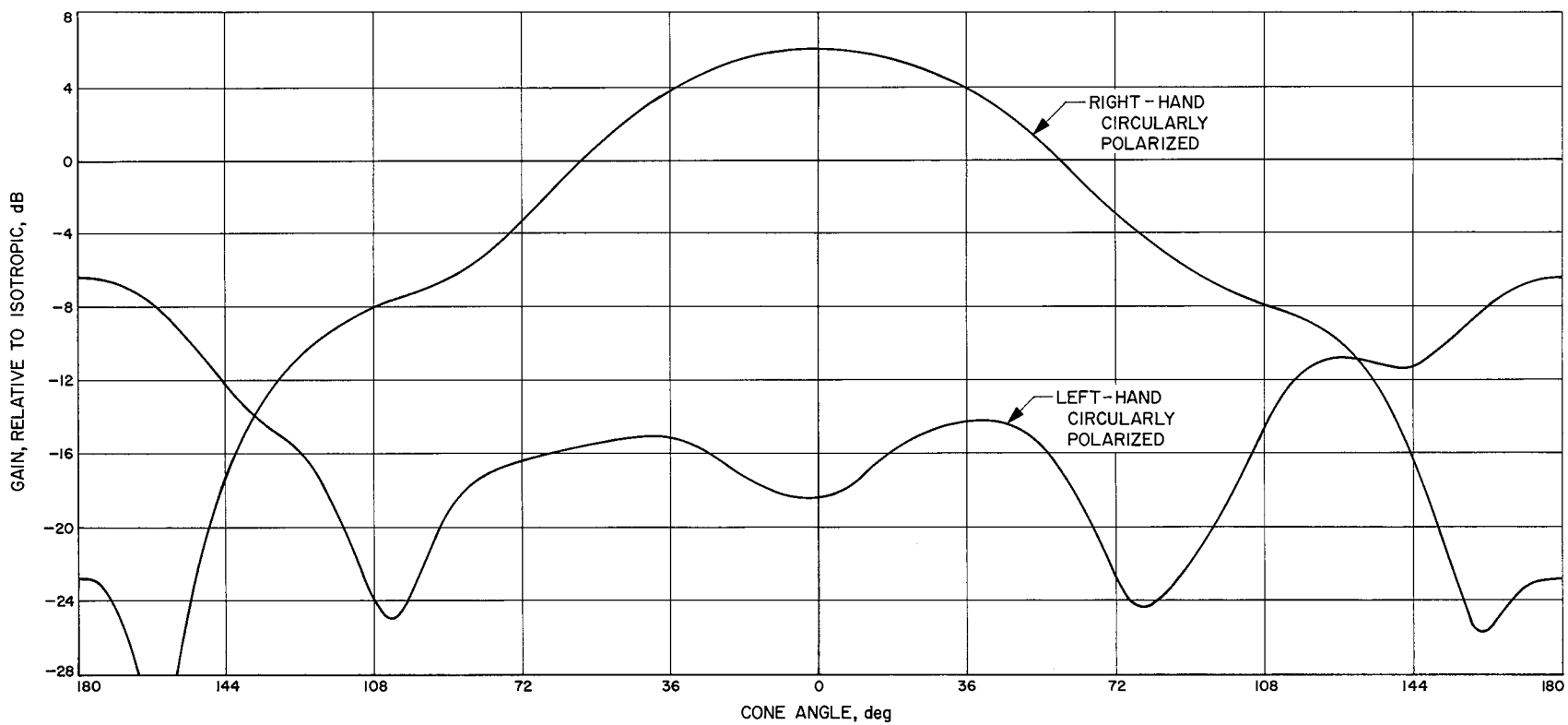


Fig. 15. Radiation patterns at 400 MHz

antenna radiates circularly polarized waves at the open end of the cavity.

c. Results. Preliminary measurements of the antenna at 400 MHz give the following electrical characteristics:

Gain	6 dB
Half-power beamwidth	84 deg
Maximum ellipticity within half-power beam	2.5 dB
VSWR at input to hybrid	1.1

The radiation patterns of the right-hand and left-hand circularly polarized components of the antenna are shown in Fig. 15. The patterns of the right-hand component at all planes are found to be substantially the same. The antenna is sterilizable in that the cavity, the probes and the connectors are all sterilizable.

Further work on the antenna will include:

- (1) Investigating power handling capability.
- (2) Reducing antenna weight by trimming the cavity walls.

Reference

1. *The Radio Amateurs Handbook by the Headquarters Staff of the American Radio Relay League*, Thirty-seventh Edition, West Hartford, Conn, pp. 90-91, 1960.

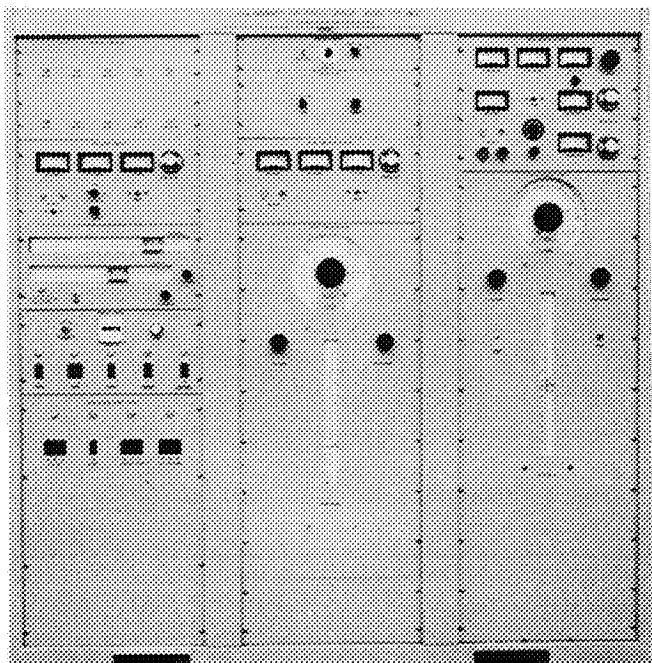


Fig. 16. 800-W CW 150-800 MHz RF power source

E. RF Breakdown Studies: Multipacting Breakdown in Coaxial Transmission Lines 150-800 MHz, R. Woo

An Eimac ETS 4800 power source (Fig. 16), which has an output of 800-W CW RF power in the 150-800 MHz

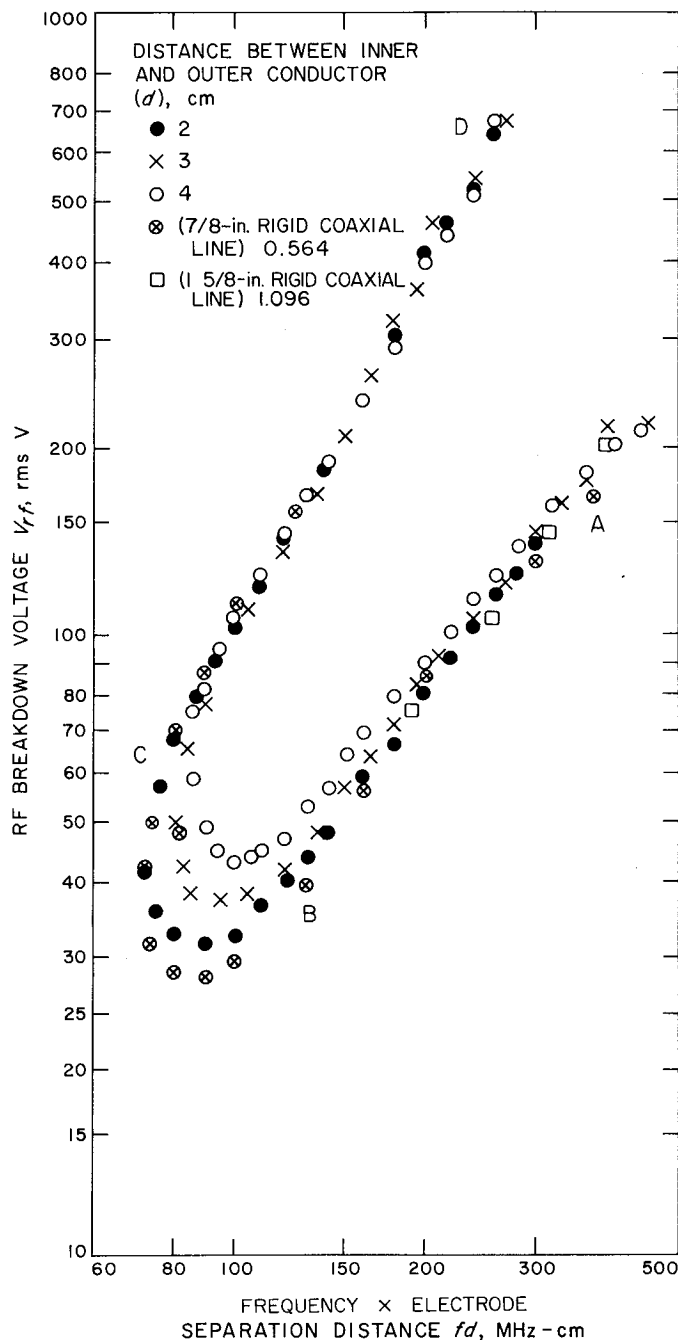


Fig. 17. Multipacting data of coaxial electrodes ($b/a = 2.3$) and $7/8$ -in. and $1 5/8$ -in. 50-ohm rigid coaxial transmission line

frequency range, was used to obtain multipacting breakdown data for $\frac{7}{8}$ -in. and 1 $\frac{5}{8}$ -in. rigid coaxial transmission line. The rigid line setup is described in SPS 37-44, Vol. IV, pp. 334-336.

The experimental results are shown in Fig. 17 along with the data for coaxial electrodes obtained in the frequency range 10-150 MHz (SPS 37-41, Vol. IV, pp. 242-246). The data show excellent scaling agreement, except for the minimum energy boundary BC where the breakdown voltages for the $\frac{7}{8}$ -in. rigid line are lower. This behavior follows the pattern of the previous data, i.e., the minimum energy boundary decreases with decreasing d . The reason is that there is a decreased electron loss to the sides for shorter inner and outer conductor separations. It should be pointed out that the effects of outgassing by

burning the discharge observed in the previous experiments at lower frequencies were also present in these experiments. For instance, boundary AB of Fig. 17 represents the threshold conditions for multipacting, but this boundary moves to higher voltages with the occurrence of multipacting. Since power is proportional to the square of voltage, this boundary rose in significant steps in terms of power, and sustaining multipacting power levels were much higher than the threshold values.

In conclusion, the above results not only serve as a further demonstration of the scaling laws for multipacting but also demonstrate that the data obtained for coaxial electrodes in SPS 37-41, Vol. IV can indeed be used for predicting multipacting breakdown in coaxial transmission line components.

PRECEDING PAGE BLANK NOT FILMED.

XXII. Spacecraft Telemetry and Command

TELECOMMUNICATIONS DIVISION

A. Multiple Mission Telemetry System: System Verification and Testing, *N. A. Burow and A. Vaisnys*

1. Introduction

The system verification and testing effort will establish a working integrated system and provide detailed performance test data on the MMTS. These data will be used to prove system concepts and to verify the results of system analysis. This work is expected to produce the final system parameters and specify the system performance for missions using the MMTS.

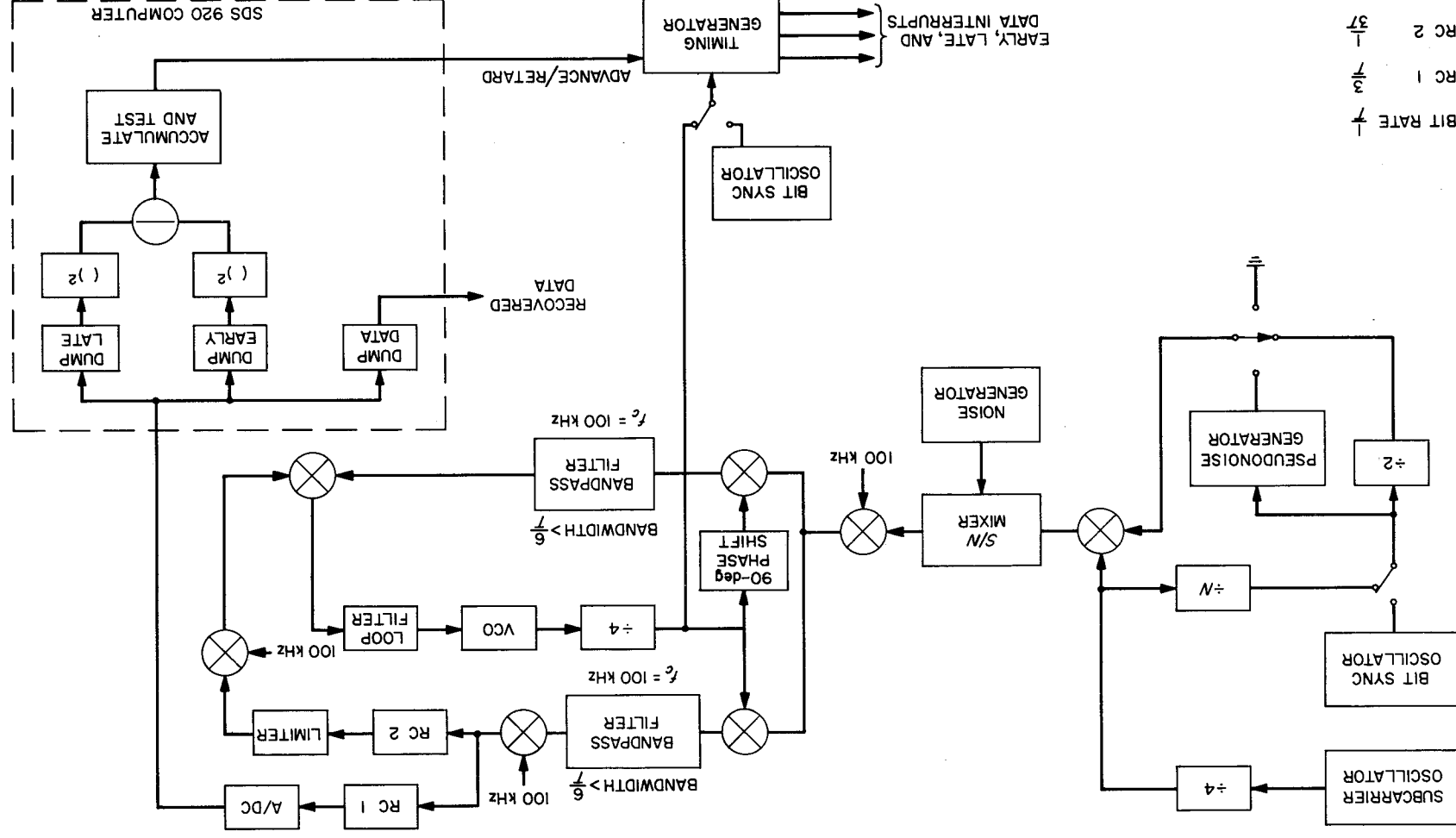
2. Preliminary Investigations

Shortly after the MMTS concept was generated, work began on the construction of a system model to demonstrate operation and to investigate the practical problems of building such a system. A Scientific Data Systems model 920 computer was available, but the RF hardware called for in the early designs was not. It was therefore necessary to model the system at audio frequencies; i.e., the 10-MHz bandpass filters were simulated by 100-kHz filters having the same bandwidth. The resulting assembly, called the baseband breadboard, very closely duplicated the functions of the actual system design except that the input signal was composed of the subcarrier(s) plus low-passed noise, instead of a phase-modulated inter-

mediate carrier plus band-passed noise. In a sense, the 10-MHz IF was replaced by 0 frequency. Figure 1 is a diagram of this system. See SPS 37-46, Vol. III, pp. 215-221 for more detailed information on the BBB system.

A prototype MMTS was completed and made available for integration and evaluation in July 1967. This system was intended to be functionally identical to the systems that will eventually go into the DSN. The prototype consists of an RF (10-MHz) subcarrier demodulator and lock detector, as well as software for the SDS 920 computer to be used in conjunction with a numerically controlled oscillator to provide bit synchronization. A system block diagram is presented in Fig. 2.

Initial setup of the prototype system was accomplished with only a few minor problems. The NCO timing generator was found to be susceptible to noise (generated primarily in the computer) causing erroneous bit sync timing. This was corrected by gating an additional timing term with the affected signals. An upconverter to convert baseband subcarrier plus data to the required 10 MHz for input to the demodulator was found to have insufficient dynamic range. Because of leakage of a 10-MHz carrier component generated by the upconverter, it was determined that the minimum input signal should be on the order of -10 dBm. The upper limit on this input signal is $+12$ dBm, thus allowing only a 22-dB dynamic range.



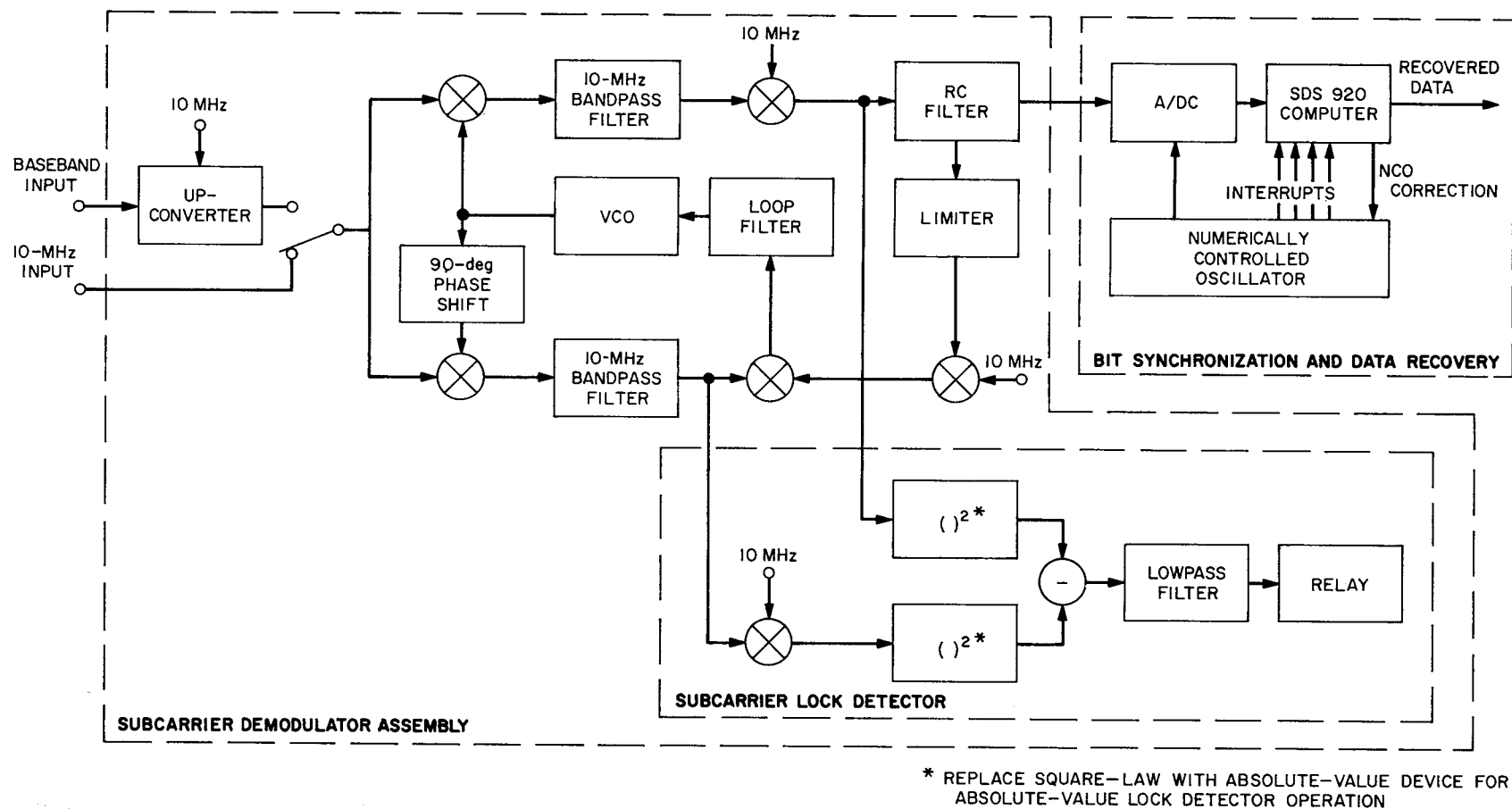


Fig. 2. Prototype MMTS block diagram

Given a subcarrier of 34.286 kHz and noise with a bandwidth of 160 kHz, the lower limit on ST/N_0 , the ratio of the signal power per bit to the noise spectral density, is +7.5 dB. To circumvent the limitation, baseband tests were performed with signal and noise passed through a bandpass filter.

3. Recent Activities

To acquire detailed performance test data on the MMTS, specific portions have been examined and evaluated in detail, including:

- (1) Ampex FR-1400 tape recorder to be used for backup recording of the demodulated subcarrier.
- (2) Square-law versus absolute-value subcarrier lock detector.
- (3) Use of an HP 4204A oscillator in the MMTS test equipment for bit sync generation and/or subcarrier simulation.
- (4) Bit synchronization jitter.
- (5) RF compatibility.

These investigations are discussed in detail below.

a. Tape recorder (Ampex FR-1400) evaluation. One of the possible backup modes of the MMT involves recording

the recovered telemetry subcarrier and later playing it back through the MMT upconverter. This mode of operation was investigated in the laboratory in order to demonstrate feasibility and to measure the degradation caused by the tape recorder. A simplified block diagram of the tape recorder test setup is shown in Fig. 3. An unmodulated subcarrier frequency was used for wow and flutter compensation, and the tape recorder was operated in the servo mode. Because of the limited dynamic range of the prototype MMT upconverter, the subcarrier and noise were passed through a bandpass filter before recording. The basic results of this investigation are as follows:

- (1) The MMTD would not operate in a satisfactory manner without some form of wow and flutter compensation.
- (2) Better compensation could be obtained when the compensation signal was recorded on the same stack of recording heads as the subcarrier. (The FR-1400 has two stacks, one for the even-numbered tracks and one for the odd-numbered tracks.)
- (3) The recording mode (AM or wide-band FM) did not have any appreciable effect on performance.

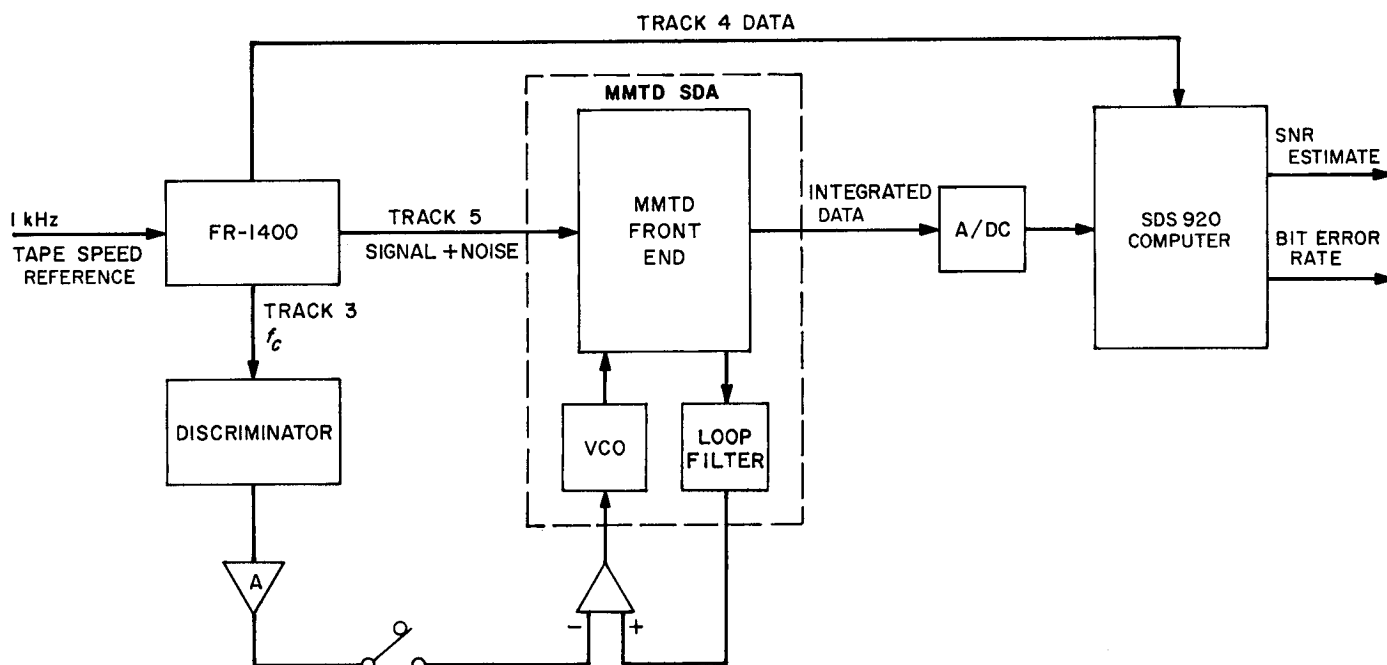


Fig. 3. Tape recorder test setup

A fairly simple wow and flutter compensation scheme was devised using an Electro-Mechanical Research Company tunable discriminator as shown in Fig. 3. Tables 1

**Table 1. Tape recorder (FR-1400) evaluation;
bit rate = 8 1/3 bits/s, subcarrier = 24.0 kHz**

Tape speed, in./s	Input ST/N_0	Minimum subcarrier $2B_{L0}$, Hz	SNORE output (effective ST/N_0)
60	No noise	0.1	+15.3
	+10.0	0.1	+ 8.5
	+ 5.0	0.1	+ 3.7
30	No noise	0.1	+17.5
	+10.0	0.1	+ 8.5
	+ 5.0	0.22	+ 3.4
15	No noise	1.0	+16.8
	+10.0	1.0	+ 8.1
	+ 5.0	10	Marginal operation ^a

^aNeither larger nor smaller subcarrier $2B_{L0}$ will work.
NOTE: Wow and flutter compensation was employed for these tests.
Bit sync $2B_{L0} = 0.3\%$.

**Table 2. Tape recorder (FR-1400) evaluation;
bit rate = 267 bits/s, subcarrier = 34.286 kHz**

Tape speed, in./s	Input ST/N_0	Minimum subcarrier $2B_{L0}$, Hz ^a	SNORE output (effective ST/N_0)
60	No noise	0.1	21.5
	+15.0	0.1	13.3
	+10.0	0.1	8.8
	+ 5.0	0.22	3.7
30	No noise	0.1	22
	+15.0	0.1	13.2
	+10.0	0.22	8.8
	+ 5.0	0.22	3.9
15 ^b	No noise	1.0	24
	+15.0	↓	13.8
	+10.0	↓	9.0
	+ 5.0	↓	4.0
7 1/2	No noise	1.0	20
	+15.0	↓	13.7
	+10.0	↓	8.7
	+ 5.0	↓	3.75

^aMinimum usable subcarrier loop bandwidth (compensated).
^bSlightly better in high speed range than low speed range.

and 2 summarize the results of the tape recorder evaluation at bit rates of 267 and 8 1/3 bits/s. The tables show minimum usable subcarrier loop bandwidth ($2B_{L0}$) and SNR degradation as a function of tape speed.

b. Subcarrier lock detector evaluation. The lock detector supplied with the prototype MMTD used square-law devices to derive an in-lock indication (system block diagram, Fig. 2). It was later proposed to substitute absolute-value circuits for the square-law circuits. An evaluation was made to determine the relative efficiency of the two configurations. The approach was to sample the output of the lock detector with an A/DC under computer control. The computer program uses the samples to compute the mean and standard deviation of the lock detector output signal and to plot its amplitude probability density. The ratio of mean to standard deviation of the output (μ/σ) is a measure of the efficiency of the lock detector and forms the basis for this comparison. Figure 4 is a plot of μ/σ vs ST/N_0 for both types of lock detectors, and does indeed indicate that the absolute value circuit approach is superior.

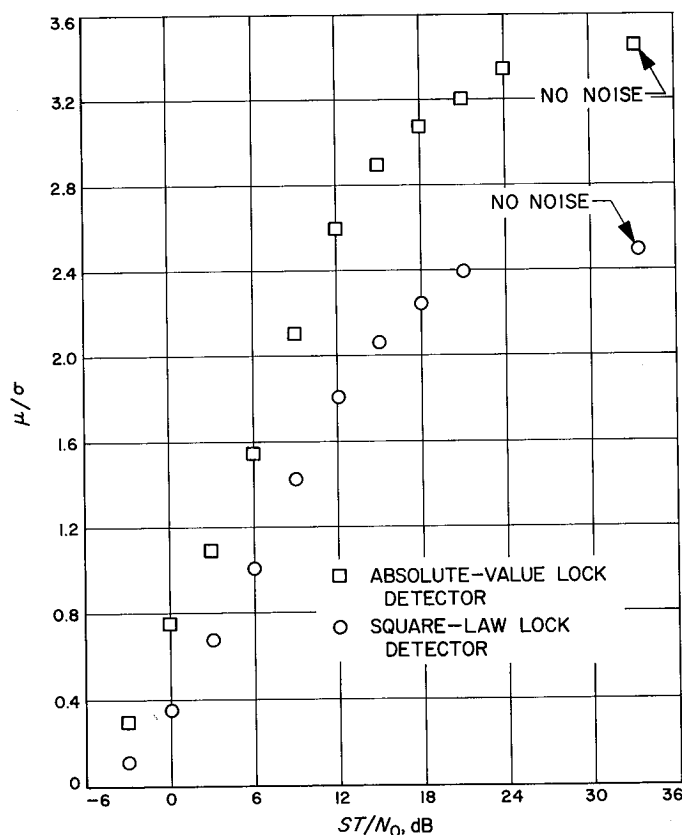


Figure 4. Comparison of square-law and absolute-value lock detectors

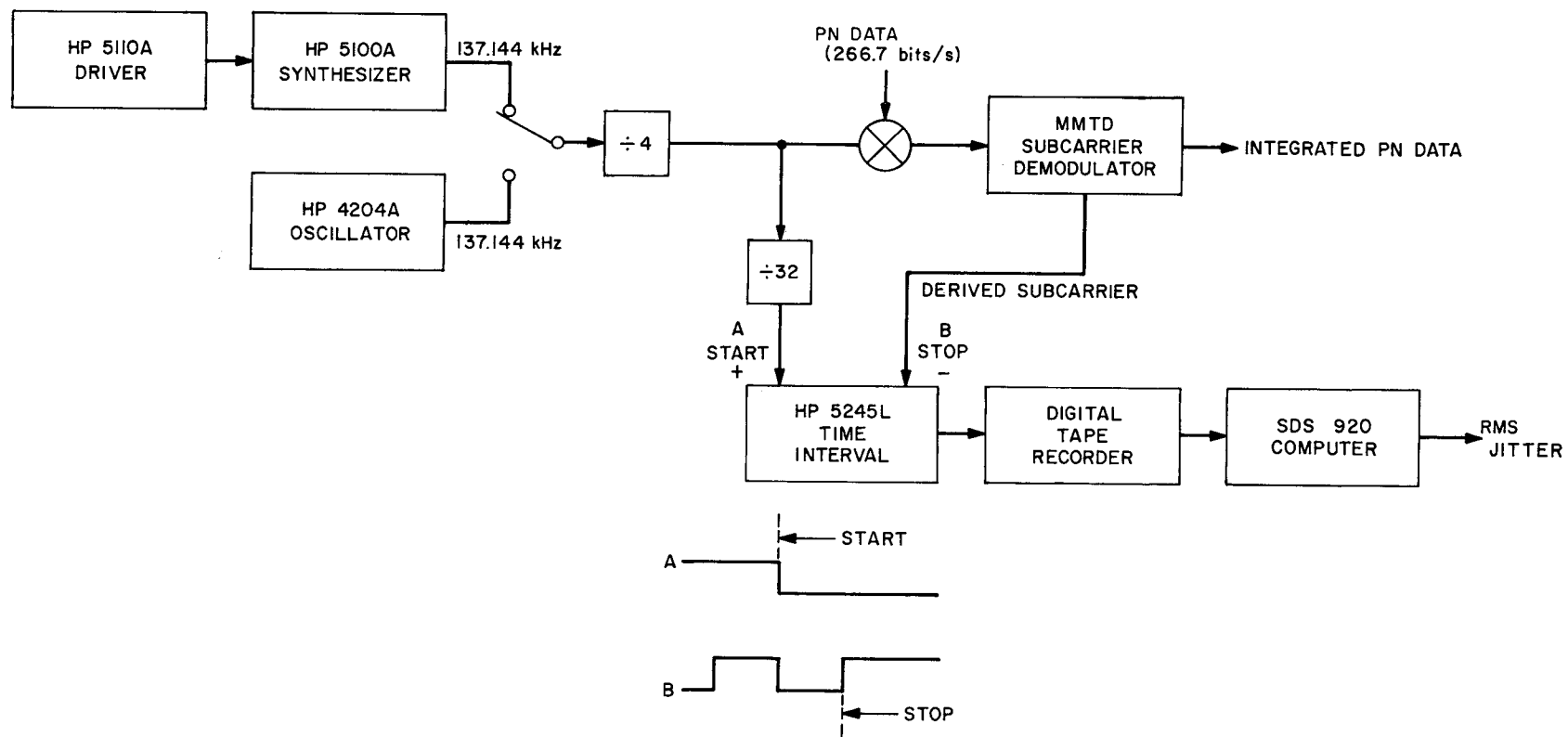


Fig. 5. Test setup for evaluation of the HP 5100A synthesizer and the HP 4204A oscillator

c. System test oscillators. Two oscillators are required as part of the MMTS test equipment to be supplied with the system. One will be used to generate the subcarrier, while the other will be used to produce bit synchronization (generally noncoherent with the subcarrier). It was desired to use a Hewlett-Packard model 4204A oscillator for both, if possible.

Jitter measurements were made to evaluate the HP 4204A for use as a bit sync oscillator. Table 3 shows data for the HP 4204A and two other oscillators. The results indicate that the type of oscillator is not as important in this application as the practice of using a higher frequency and dividing down to gain stability.

Table 3. Oscillator jitter measurements

Oscillator	Configuration	σ , μ s rms
HP 4204A	10 Hz \div 1	13.1
HP 5100A/5110	10 Hz \div 1	21.2
HP 4204A	100 Hz \div 1	1.8
HP 5100A/5110	100 Hz \div 1	2.1
HP 241A	100 Hz \div 1	2.3
HP 4204A	10.24 kHz \div 1024	0.08
HP 5100A/5110	10.24 kHz \div 1024	0.04
HP 4204A	102.4 kHz \div 1024	0.04
HP 5100A/5110	102.4 kHz \div 1024	0.03

4204A—Inexpensive new audio-oscillator
5100A/5110—Synthesizer and driver
241A—Pushbutton audio-oscillator

The HP 4204A was also evaluated for use as a subcarrier simulator. An HP model 5100A synthesizer was used as a reference. The test setup is shown in Fig. 5. Root mean square phase jitter between the transmitted and derived subcarriers versus subcarrier loop bandwidth $2B_{L0}$ was measured. The test was run at no noise with 50,000 data samples recorded for each point. The results are plotted in Fig. 6. Overall system performance (again at no noise) for both the HP 4204A oscillator and the HP 5100A synthesizer was measured using the test configuration of Fig. 7. The results are plotted in Fig. 8. Both tests indicate that the HP 4204A has too much phase jitter for use as a subcarrier simulator for the MMTS.

d. Computer bit synchronization. Measurement of bit sync jitter versus ST/N_0 and loop bandwidth (in percent of the bit period) was made for loop algorithms identified as "first order" and "second order." The data are presented

in Table 4. A hunting effect due to insufficient resolution in the second-order loop's error accumulator was determined to be causing the excessive jitter seen in the second-order loop. A solution for this problem has been found, and the second-order system is being re-evaluated.

e. RF compatibility testing. Tests were performed on the MMTS over an S-band RF link using a DSN receiver and an S-band transmitter. These tests were intended to show compatibility between the MMTS and the *Mariner*

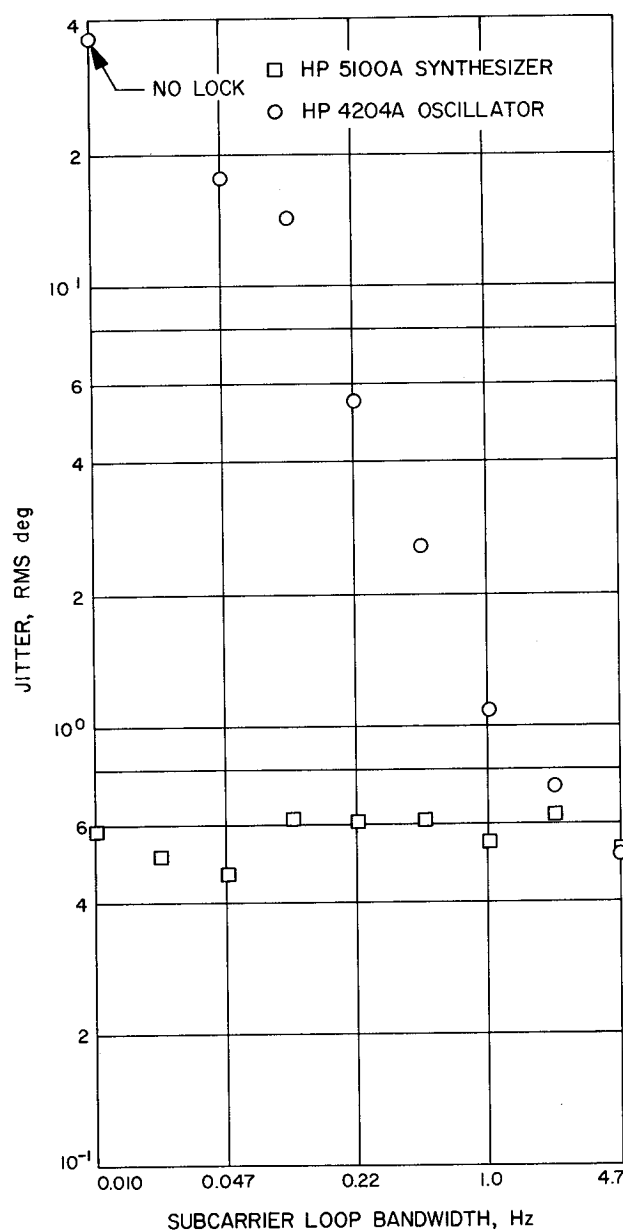


Fig. 6. Performance comparison of the HP 5100A synthesizer and the HP 4204A oscillator

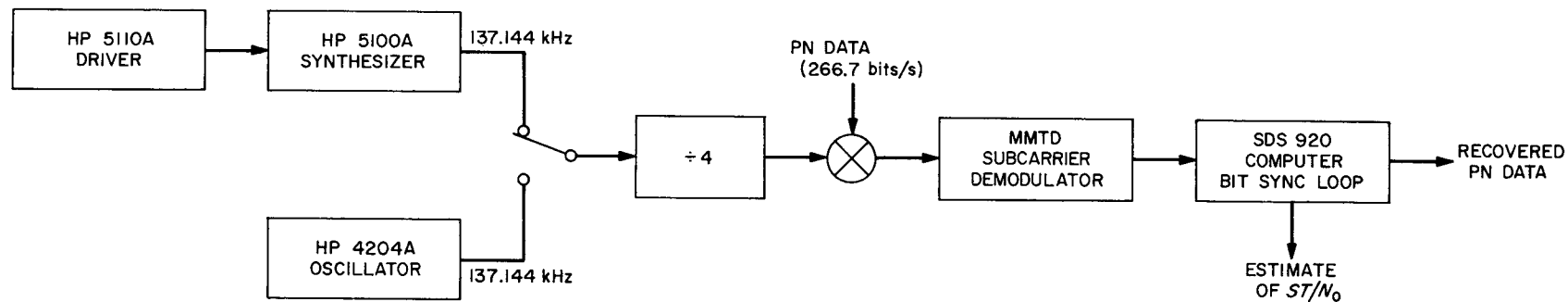


Fig. 7. Test setup to compare overall system performance using either the HP 5100A synthesizer or the HP 4204A oscillator

**Table 4. RMS bit sync jitter; bit rate = 266.7 bits/s,
bandwidth = 0.1 %**

ST/N ₀ , dB	1st-order loop jitter		2nd-order loop jitter	
	μs	Deg	μs	Deg
No noise	4.5	0.43	68.0	6.54
+10.0	11.8	1.13	72.2	6.95
+ 5.0	27.2	2.61	99.7	9.6
0.0	68.0	6.52	486.9	48.0

Mars 1969 flight telemetry system. Additional objectives were to measure the relative performance of the FTS to a laboratory standard signal source, to investigate the feasibility of using a 3-Hz DSN receiver loop, and to obtain additional data on radio loss. Tests were completed for telemetry channels A and B alone. The results are currently being analyzed by the *Mariner* Mars 1969 project. Further testing using multiplexed combinations of channels A, B, and C are planned for early 1968.

4. Future Investigation

At present, the subcarrier demodulator assembly of the MMTS is being updated to the final DSN configuration. An updated version of the data conditioner, which is a part of the SDA, has already been received and is being evaluated using baseband data and noise. Among the areas of investigation are the following:

- (1) Bit sync jitter and performance degradation as a function of bit sync loop bandwidth.
- (2) Performance comparison of first- and second-order bit sync loops.
- (3) Effect on performance of dc offsets in the data.
- (4) Effectiveness of a dc blocking scheme.

When the updated SDA is returned, overall system performance will be evaluated, including a comparison of actual to theoretical losses in both the subcarrier and bit sync loops.

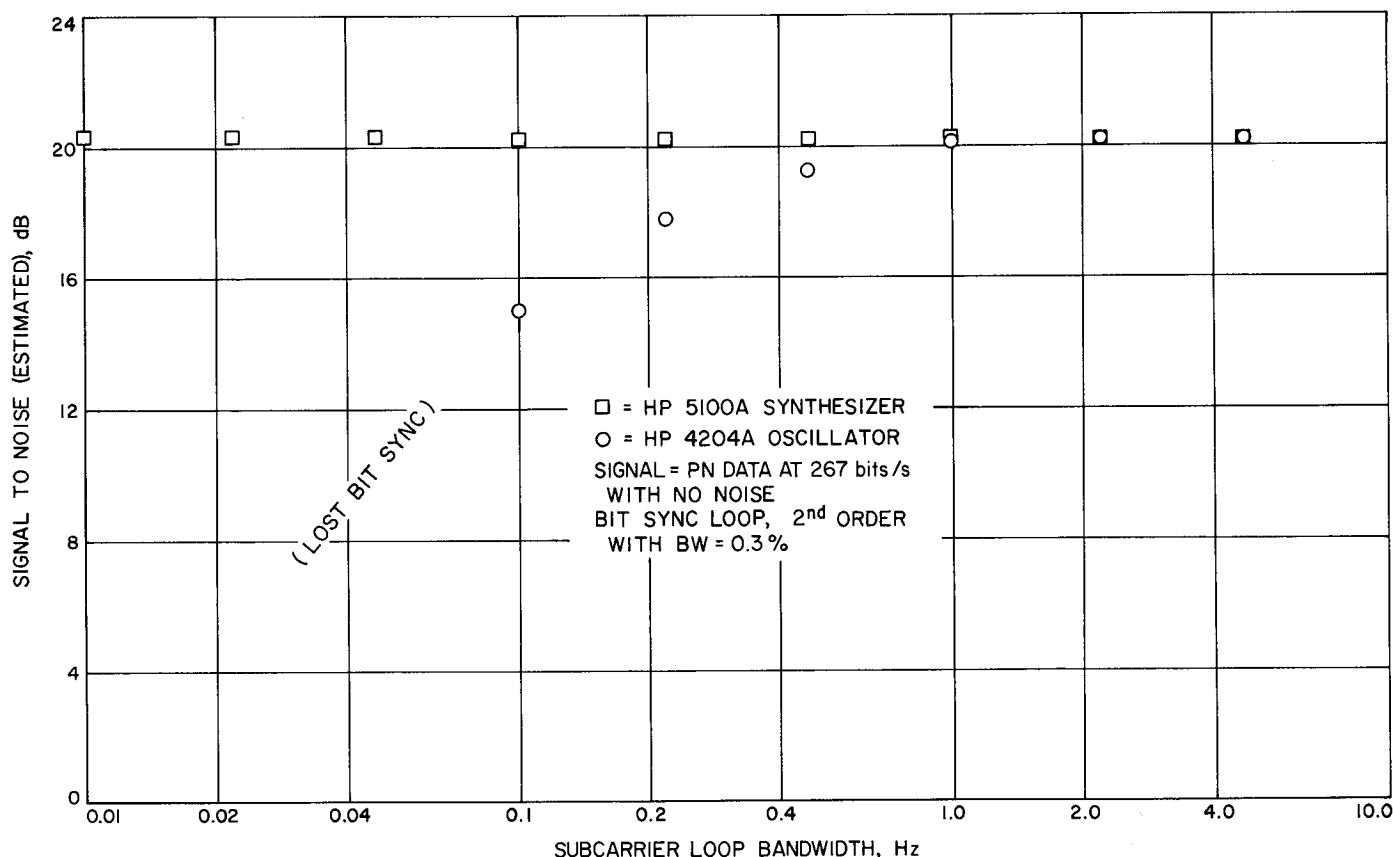


Fig. 8. Comparison of system performance using the HP 5100A synthesizer and the HP 4204A oscillator

B. Time Synchronization in an MFSK Receiver,

H. D. Chadwick

1. Introduction

An incoherent m -ary frequency shift keying communication system is under consideration for possible future planetary probes. The advantage of such a system lies in its incoherent operation, which permits operation at signal-to-noise ratios below the threshold of phase lock systems. One optimum detector for an MFSK link has been shown to be the spectrum analyzer receiver,¹ in which the power spectrum of the received waveform is calculated and the peak frequency is chosen as the most probable transmitted frequency.

An estimate of the power density spectrum of the input signal may be obtained either by direct Fourier transformation of a T -second sample, where the squared magnitude of the Fourier transformation serves as the estimated power spectrum, or by finding the autocorrelation function of the T -second input sample and taking the Fourier transform of the autocorrelation function. Ideally, both techniques are equivalent, but the autocorrelation function technique permits truncation, or "hanning" in the time domain to reduce variations in the

estimated power spectrum due to frequency instabilities in the signals (Ref. 1 and SPS 37-33, Vol. III, pp. 103-107). However, for both techniques, optimum operation (suboptimum operation in the truncated case) requires accurate synchronization in the starting time (epoch) of each word.

In this article, a special case of the time synchronization problem is analyzed, and a technique is described for automatic time synchronizing in an MFSK receiver. The analysis is based on the following major assumptions:

- (1) Perfect frequency synchronization is assured.
- (2) A known synchronization sequence can be transmitted ahead of the data.
- (3) The transmitter and receiver are stable enough so that once synchronization is obtained it remains for the entire data transmission.

2. Error Due to Incorrect Time Synchronization

The probability of error for an MFSK receiver in perfect frequency and time synchronization has been shown (Ref. 2) to be

$$P_e = 1 - \int_0^\infty x \exp \left[-\frac{1}{2}(x^2 + \gamma^2) \right] I_0(\gamma x) [1 - \exp(-x^2/2)]^{m-1} dx \quad (1)$$

where

m = the number of signals in the signaling set

$\gamma^2 = \frac{E}{N_0}$ = signal-to-noise spectral density ratio

N_0 = two-sided noise spectral density

E = signal energy per word

$I_0(\)$ = modified Bessel function of the first kind, order zero

In Section 2a below, a similar expression is derived for a spectrum analyzer receiver in arbitrary time synchronization (assuming correct frequency synchronization) which reduces to Eq. (1) when the time synchronization is correct.

The spectrum analyzer receiver assumed in this report is based on the use of the fast Fourier transform to calculate the received spectrum. Although the advantage of the autocorrelation function technique is that truncation may be performed in the time domain to reduce the effects of frequency instability¹ (Ref. 1 and SPS 37-33, Vol. III, pp. 103-107) and to gain speed in calculation, the use of the direct Fourier transform technique permits greatly simplified analysis, and the equivalent truncation may also be performed in the frequency domain if desired. Perfect frequency synchronization has been assumed in this report; consequently, truncation processing has not been included, and the direct Fourier transform technique is used.

a. Receiver model. The transmitted signal is assumed to be of the form:

$$s_j(t) = A \cos(2\pi f_i t), \quad \left. \begin{array}{l} jT \leq t \leq (j+1)T \\ i = 1, 2, \dots, m \end{array} \right\} \quad (2)$$

¹Charles, F., and Shein, N., "A Preliminary Study of the Application of Noncoherent Techniques to Low Power Telemetry," JPL Section 334 internal memorandum, Nov. 15, 1965.

where the m frequencies f_i are such that $f_i = k/T$, where k is an integer.

At the receiver, the signal is corrupted by additive white Gaussian noise, with noise spectral density N_0 watts/Hz. The received signal is represented by

$$x(t) = A \cos(2\pi f_i t + \phi) + n(t) \quad (3)$$

where ϕ is an unknown, uniformly distributed phase term due to the incoherent nature of the detection. The receiver samples the incoming waveform at a rate of N/T samples per second.

It is assumed that the receiver bandwidth is W Hz and that the sampling is at the Nyquist rate. Therefore,

$$\frac{N}{T} = 2W \quad (4)$$

The samples

$$x_i = x(t_i) = x\left(\frac{i}{N} T\right), \quad i = 0, 1 \dots N-1$$

are statistically independent Gaussian random variables with mean $S(t_i)$ and variance

$$\sigma_x^2 = 2N_0 W = N_0 \frac{N}{T}$$

The spectrum analyzer receiver then calculates (preferably by the fast Fourier transform technique) the discrete Fourier coefficients of the samples x_i at each of the m possible transmitted frequencies. The discrete Fourier coefficients at frequency f_j are given by (Ref. 3)

$$a_j = \sum_{i=0}^{N-1} x_i \cos\left(2\pi f_j \frac{i}{N} T\right) \quad (5)$$

$$b_j = \sum_{i=0}^{N-1} x_i \sin\left(2\pi f_j \frac{i}{N} T\right) \quad (6)$$

and the magnitude of the spectral component is

$$r_j = (a_j^2 + b_j^2)^{1/2} \quad (7)$$

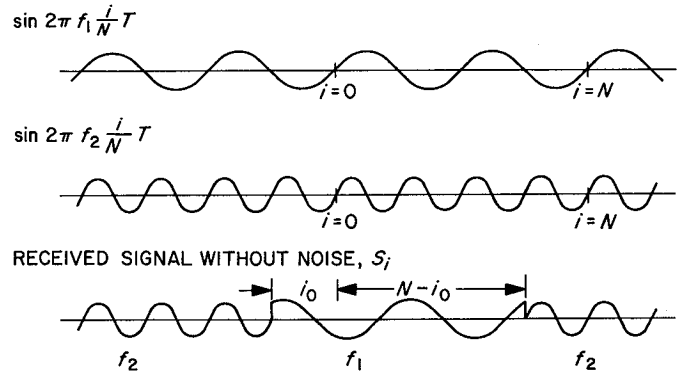


Fig. 9. Timing diagram of the Fourier transform process

If the receiver is not time-synchronized with the transmitter, the N samples will overlap two frequencies (Fig. 9).² When $N - i_0$ samples are taken of the first frequency f_1 and i_0 samples of the second frequency f_2 , then the overlap coefficient ρ is defined:

$$\rho \triangleq \frac{i_0}{N} \quad (8)$$

For $\rho = 0$, the receiver is in perfect synchronization. For $\rho > 0$, there will be contributions to the Fourier coefficients at the two frequencies f_1 and f_2 . These components are determined by

$$a_1 = \sum_{i=0}^{N-1} x_i \cos\left(2\pi f_1 \frac{i}{N} T\right) \quad (9)$$

$$b_1 = \sum_{i=0}^{N-1} x_i \sin\left(2\pi f_1 \frac{i}{N} T\right) \quad (10)$$

$$a_2 = \sum_{i=0}^{N-1} x_i \cos\left(2\pi f_2 \frac{i}{N} T\right) \quad (11)$$

$$b_2 = \sum_{i=0}^{N-1} x_i \sin\left(2\pi f_2 \frac{i}{N} T\right) \quad (12)$$

²The time reference, $t = 0$, or equivalently, $i = 0$, is taken always to be the start of the summation for the discrete Fourier transform.

The Fourier components a_1, b_1, a_2, b_2 will be Gaussian-distributed because the discrete Fourier transform is a linear process. The samples x_i between the indices 0 and $N-1$ will be

$$x_i = \begin{cases} A \cos \left(2\pi f_1 \frac{(i + i_0)}{N} T + \phi_1 \right) + n_1, & 0 \leq i \leq N - i_0 - 1 \\ A \cos \left(2\pi f_2 \frac{(i + i_0)}{N} T + \phi_2 \right) + n_1, & N - i_0 \leq i \leq N - 1 \end{cases} \quad (13)$$

In calculating the expected value of the Fourier components, the zero mean noise term may be ignored. Therefore,

$$E(a_1) = \sum_{i=0}^{N-i_0-1} A \cos \left[2\pi f_1 \frac{(i + i_0)}{N} T + \phi_1 \right] \cos \left(2\pi f_1 \frac{i}{N} T \right) + \sum_{i=N-i_0}^{N-1} A \cos \left[2\pi f_2 \frac{(i + i_0)}{N} T + \phi_2 \right] \cos \left(2\pi f_2 \frac{i}{N} T \right) \quad (14)$$

If it is assumed that $(i_0/N)T$ is an integer multiple of the period of the difference frequency $1/(f_1 - f_2)$, then the equation above may be simplified greatly:³

$$E(a_1) = A \cos \phi_1 \sum_{i=0}^{N-i_0-1} \cos^2 \left(2\pi f_1 \frac{i}{N} T \right) \quad (15)$$

Because the second term involves the summation over an integral number of periods of the product of two different frequency cosine waves and because the index of the first summation may be shifted an integral number of periods without changing the summation, Eq. (15) reduces to⁴

$$E(a_1) = A \left(\frac{N - i_0}{2} \right) \cos \phi_1 \quad (16)$$

Similarly, the remaining expected values can be shown to be

$$E(b_1) = -A \left(\frac{N - i_0}{2} \right) \sin \phi_1 \quad (17)$$

$$E(a_2) = A \left(\frac{i_0}{2} \right) \cos \phi_2 \quad (18)$$

$$E(b_2) = -A \left(\frac{i_0}{2} \right) \sin \phi_2 \quad (19)$$

The variance of the Fourier coefficient can also be found easily. Leaving out the fixed terms

$$\sigma_1^2 = \text{var } a_1 = E \left[\sum_{i=0}^{N-1} n_i \cos \left(2\pi f_1 \frac{i}{N} T \right) \sum_{j=0}^{N-1} n_j \cos \left(2\pi f_1 \frac{j}{N} T \right) \right] \quad (20)$$

³This assumption may be more easily justified if the difference frequency $f_1 - f_2$ is large.

⁴See Jolley (Eq. 438, Ref. 4) for large N .

which can be written

$$\sigma_1^2 = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} E(n_i n_j) \cos \left(2\pi f_1 \frac{i}{N} T \right) \cos \left(2\pi f_2 \frac{j}{N} T \right) \quad (21)$$

the independence of the noise samples gives

$$E(n_i n_j) = N_0 \frac{N}{T} \delta_{ij} \quad (22)$$

with the result that

$$\sigma_1^2 = \sum_{i=0}^{N-1} N_0 \frac{N}{T} \cos^2 \left(2\pi f_1 \frac{i}{N} T \right) \quad (23)$$

and, since the result is the same with all the Fourier coefficients,

$$\sigma^2 = \frac{N_0}{2T} N^2 \quad (24)$$

By making the substitutions

$$B = \frac{AN}{2}$$

and

$$\epsilon = \rho - \frac{1}{2} = \frac{i_0}{N} - \frac{1}{2}$$

the probability density functions of the Fourier coefficients can be written

$$p(a_1 | \phi_1) = \frac{1}{(2\pi)^{1/2} \sigma} \exp \left\{ -\frac{1}{2\sigma^2} \left[a_1 - B \left(\frac{1}{2} - \epsilon \right) \cos \phi_1 \right]^2 \right\} \quad (25)$$

$$p(b_1 | \phi_1) = \frac{1}{(2\pi)^{1/2} \sigma} \exp \left\{ -\frac{1}{2\sigma^2} \left[b_1 + B \left(\frac{1}{2} - \epsilon \right) \sin \phi_1 \right]^2 \right\} \quad (26)$$

$$p(a_2 | \phi_2) = \frac{1}{(2\pi)^{1/2} \sigma} \exp \left\{ -\frac{1}{2\sigma^2} \left[a_2 - B \left(\frac{1}{2} + \epsilon \right) \cos \phi_2 \right]^2 \right\} \quad (27)$$

$$p(b_2 | \phi_2) = \frac{1}{(2\pi)^{1/2} \sigma} \exp \left\{ -\frac{1}{2\sigma^2} \left[b_2 + B \left(\frac{1}{2} + \epsilon \right) \sin \phi_2 \right]^2 \right\} \quad (28)$$

The probability densities of the variables

$$r_1 = (a_1^2 + b_1^2)^{1/2}$$

and

$$r_2 = (a_2^2 + b_2^2)^{1/2}$$

are found by forming the joint densities, converting to polar coordinates, and integrating out the uniformly distributed phase parameters ϕ_1 and ϕ_2 . The resulting densities are

$$p(r_1) = \frac{r_1}{\sigma^2} \exp \left\{ -\frac{1}{2\sigma^2} \left[r_1^2 + B^2 \left(\frac{1}{2} - \epsilon \right)^2 \right] \right\} I_0 \left[\frac{B \left(\frac{1}{2} - \epsilon \right) r_1}{\sigma^2} \right] \quad (29)$$

$$p(r_2) = \frac{r_2}{\sigma^2} \exp \left\{ -\frac{1}{2\sigma^2} \left[r_2^2 + B^2 \left(\frac{1}{2} + \epsilon \right)^2 \right] \right\} I_0 \left[\frac{B \left(\frac{1}{2} + \epsilon \right) r_2}{\sigma^2} \right] \quad (30)$$

In an m -ary detector ($m > 2$) the remaining spectral terms (at frequencies other than f_1 and f_2) will have the Rayleigh distributions

$$p(r_k) = \frac{r_k}{\sigma^2} \exp \left(-\frac{r_k^2}{2\sigma^2} \right), \quad 2 \leq k \leq m \quad (31)$$

A correct decision is made by the m -ary detector if the component $r_1 > r_j$ for all j ($2 \leq j \leq m$). The probability of a correct decision is thus given by

$$\begin{aligned} P_c &= 1 - P_e = P(r_1 > r_j; \text{all } j \neq 1) \\ &= P(r_j < r_1; \text{all } j \neq 1) \end{aligned} \quad (32)$$

Since the r_j values are statistically independent, this may be written

$$\begin{aligned} P_c &= \int_0^\infty P(r_2 < x, r_3 < x, \dots, r_m < x | r_1 = x) P(r_1 = x) dx \\ &= \int_0^\infty P(r_2 < x) P(r_3 < x) \dots P(r_m < x) P(r_1 = x) dx \end{aligned} \quad (33)$$

From the results above,

$$\begin{aligned} P(r_2 < x) &= \int_0^x P(r_2) dr_2 \\ &= \int_0^x \frac{r_2}{\sigma^2} \exp \left\{ -\frac{1}{2} \left[\frac{r_2^2}{\sigma^2} + \gamma^2 \left(\frac{1}{2} + \epsilon \right)^2 \right] \right\} I_0 \left[\frac{\gamma \left(\frac{1}{2} + \epsilon \right) r_2}{\sigma} \right] dr_2 \\ &= \int_0^{x/\sigma} y \exp \left\{ -\frac{1}{2} \left[y^2 + \gamma^2 \left(\frac{1}{2} + \epsilon \right)^2 \right] \right\} I_0 \left[\gamma \left(\frac{1}{2} + \epsilon \right) y \right] dy \end{aligned} \quad (34)$$

using the substitutions

$$\gamma^2 = \frac{E}{N_0} = \frac{B^2}{\sigma^2}$$

for

$$\frac{N}{T} = 2W$$

and

$$y = \frac{r_2}{\sigma}$$

Using Marcum's Q Function (Ref. 5), which is defined by

$$Q(\alpha, \beta) = \int_{\beta}^{\infty} x \exp \left[- \left(\frac{x^2 + \alpha^2}{2} \right) \right] I_0(\alpha x) dx \quad (35)$$

where

$$Q(\alpha, 0) = 1$$

$$Q(0, \beta) = \exp \left(- \frac{\beta^2}{2} \right)$$

the expression for $P(r_2 < x)$ can be written

$$P(r_2 < x) = 1 - Q \left[\gamma \left(\frac{1}{2} + \epsilon \right), \frac{x}{\sigma} \right] \quad (36)$$

For the remaining $m - 2$ components,

$$\begin{aligned} P(r_j < x) &= \int_0^x \frac{r_j}{\sigma^2} \exp \left(\frac{r_j^2}{2\sigma^2} \right) dr_j \\ &= 1 - \exp \left(- \frac{x^2}{2\sigma^2} \right) \end{aligned} \quad (37)$$

The complete expression for the probability of error as a function of the timing uncertainty can thus be written

$$P_e = 1 - P_c =$$

$$1 - \int_0^{\infty} \left\{ 1 - Q \left[\gamma \left(\frac{1}{2} + \epsilon \right), \frac{x}{\sigma} \right] \right\} \left[1 - \exp \left(\frac{x^2}{2\sigma^2} \right) \right]^{m-2} \frac{x}{\sigma^2} \exp \left\{ - \frac{1}{2} \left[\frac{x^2}{\sigma^2} + \gamma^2 \left(\frac{1}{2} - \epsilon \right)^2 \right] \right\} I_0 \left[\frac{\gamma \left(\frac{1}{2} - \epsilon \right) x}{\sigma} \right] dx \quad (38)$$

or, by the substitution $y = x/\sigma$,

$$P_e =$$

$$1 - \int_0^{\infty} \left\{ 1 - Q \left[\gamma \left(\frac{1}{2} + \epsilon \right), y \right] \right\} \left[1 - \exp \left(- \frac{y^2}{2} \right) \right]^{m-2} y \exp \left\{ - \frac{1}{2} \left[y^2 + \gamma^2 \left(\frac{1}{2} - \epsilon \right)^2 \right] \right\} I_0 \left[\gamma \left(\frac{1}{2} - \epsilon \right) y \right] dy \quad (39)$$

It can be seen that when $\epsilon = -1/2$ (no timing error), the term

$$Q\left[\gamma\left(\frac{1}{2} + \epsilon\right), y\right]$$

becomes

$$\exp\left(-\frac{y^2}{2}\right)$$

and Eq. (39) reduces to that of Arthurs and Dym (Eq. 1).

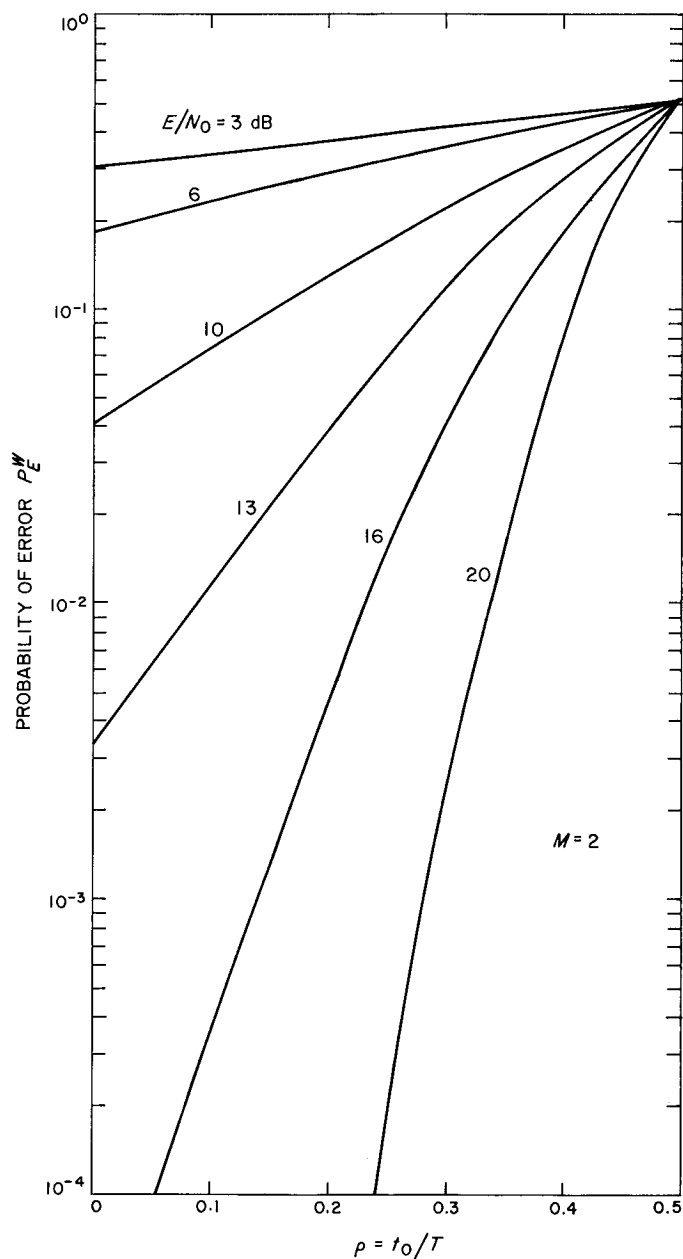


Fig. 10. Probability of error per word vs time displacement for fixed time displacement, $m = 2$

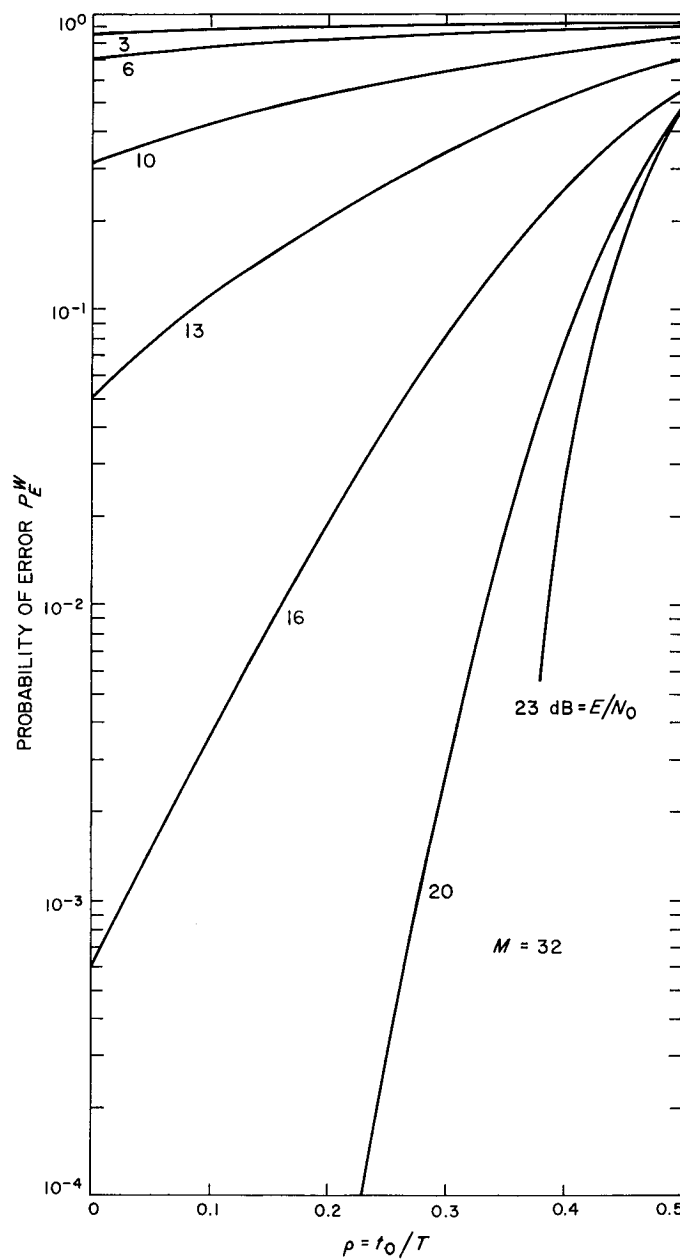


Fig. 11. Probability of error per word vs time displacement for fixed time displacement, $m = 32$

b. Results. The probability of error from Eq. (39) has been evaluated by numerical integration on a digital computer. The results are shown in Figs. 10 and 11 for $m = 2$ and $m = 32$. The figures illustrate that a small timing uncertainty can result in considerable increase in the probability of error. Since these results are based on the assumptions described in the introduction, they are not completely representative of the results that would be obtained in a more realistic system. The effect of including such additional factors as, for example, frequency uncertainty, would be only to degrade the performance of the system. These results can therefore be regarded as a lower bound to the probability of error to be encountered in an actual system.

3. Maximum Likelihood Estimator for Timing Error

To minimize the error due to timing uncertainty, it would be desirable to estimate the time shift between

the received waveform and the receiver time origin and to use this estimate to correct the position of the receiver time origin. One method of performing this estimate is to use the maximum likelihood estimator $\hat{\epsilon}$ of the percentage time shift ϵ .

Since the input samples x_i by themselves contain no information about the quantity ϵ , an estimate can be performed only after the discrete Fourier transform process. To simplify the computation it is assumed that a known synchronization sequence, f_1 followed by f_2 , is transmitted and repeated as often as necessary. Then it is required to estimate the time of transition from f_1 to f_2 relative to the receiver's present time origin, or, equivalently, to estimate the parameter $\hat{\epsilon}$.

The joint density function of the independent statistics a_1 , b_1 , a_2 , and b_2 can be written by using Eqs. (25-28):

$$p(a_1, b_1, a_2, b_2) = p(a_1) p(b_1) p(a_2) p(b_2)$$

$$= \left[\frac{1}{(2\pi\sigma)^{1/2}} \right] \exp - \frac{1}{2\sigma^2} \left\{ \left[a_1 - B \left(\frac{1}{2} - \epsilon \right) \cos \phi_1 \right]^2 + \left[b_1 + B \left(\frac{1}{2} - \epsilon \right) \sin \phi_1 \right]^2 \right.$$

$$\left. + \left[a_2 - B \left(\frac{1}{2} + \epsilon \right) \cos \phi_2 \right]^2 + \left[b_2 + B \left(\frac{1}{2} - \epsilon \right) \sin \phi_2 \right]^2 \right\} \quad (40)$$

To determine the maximum likelihood estimator for an unknown parameter, the expression for $p(a_1, b_1, a_2, b_2)$, or equivalently for $\log p(a_1, b_1, a_2, b_2)$, is maximized with respect to the parameter by differentiating and setting the derivative equal to zero. In addition to the parameter ϵ , the parameters ϕ_1 , ϕ_2 , and B are also unknown and must be eliminated from the expression for $\hat{\epsilon}$. This is done by solving also for $\hat{\phi}_1$, $\hat{\phi}_2$, and \hat{B} and substituting these relations into the expression for $\hat{\epsilon}$.

This derivation of the maximum likelihood estimator is performed by standard techniques.⁵ The resulting expression is

$$\hat{\epsilon} = \frac{r_2 - r_1}{2(r_1 + r_2)} \quad (41)$$

a. The distribution of $\hat{\epsilon}$. The probability density function of $\hat{\epsilon}$ can be derived from the densities $p(r_1)$ and $p(r_2)$:

$$p(\hat{\epsilon}) = \left(\frac{1}{4} - \hat{\epsilon}^2 \right) \exp \left[- \frac{\gamma^2}{2} \left(\epsilon^2 + \frac{1}{2} \right) \right]$$

$$\times \int_0^\infty y^3 \exp \left[- \frac{y^2}{2} \left(\hat{\epsilon}^2 + \frac{1}{2} \right) \right] I_0 \left[\gamma \left(\frac{1}{2} - \epsilon \right) \left(\frac{1}{2} - \hat{\epsilon} \right) y \right] I_0 \left[\gamma \left(\frac{1}{2} + \epsilon \right) \left(\frac{1}{2} + \hat{\epsilon} \right) y \right] dy \quad (42)$$

⁵Complete derivations are given in Chadwick, H., "Time Synchronization in an MFSK Receiver," JPL Section 334 internal memorandum, Nov. 1967.

This is a formidable expression, which so far has resisted all efforts to put it in closed form. Numerical integration on the computer for different values of the parameters ϵ and α has led to some description of its behavior, however. Figure 12 shows one of these numerically integrated functions. It can be seen from these curves that the distribution is symmetrical for $\epsilon = 0$ and becomes biased for $\epsilon \neq 0$. As would be expected, the variance of $\hat{\epsilon}$ increases with decreasing signal-to-noise ratio γ^2 . The variance of the estimate $\hat{\epsilon}$ versus the signal-to-noise ratio is plotted in Fig. 13. The degree of bias $E(\hat{\epsilon})$ versus ϵ is shown in Fig. 14.

b. Experimental results. An experiment was devised to verify the probability densities $p(\hat{\epsilon})$ for different values of the parameters ϵ and γ^2 . The experimental setup is shown in Fig. 15. The two frequencies were obtained by switching a frequency synthesizer back and forth at a rate of 4 times per second. The signal was mixed with

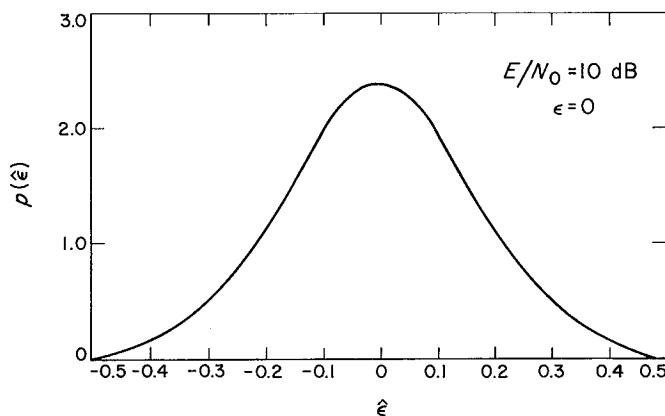


Fig. 12. Predicted probability density of time shift estimator $P(\hat{\epsilon})$, $E/N_0 = 10$ dB

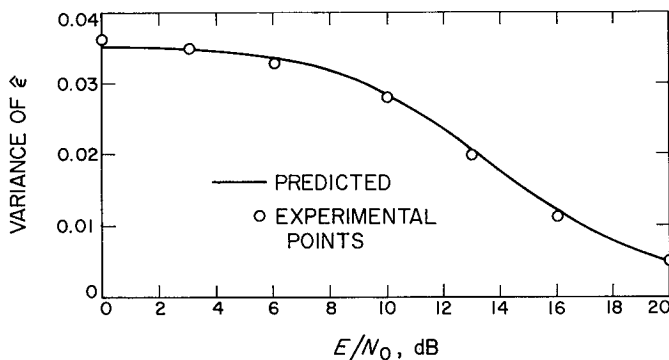


Fig. 13. Variance of time shift estimator $\hat{\epsilon}$ vs signal-to-noise ratio

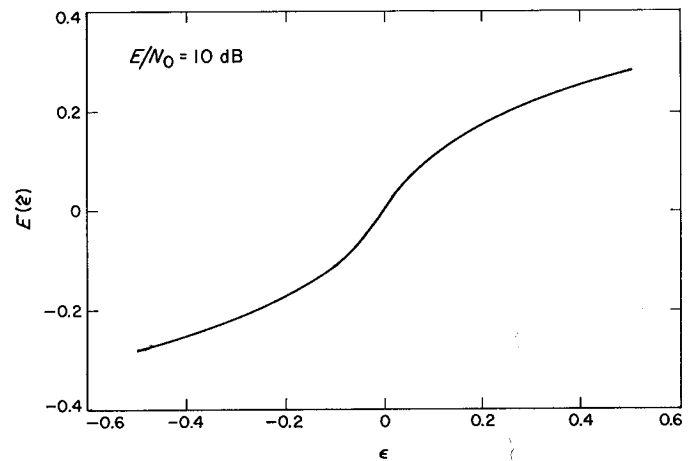


Fig. 14. Bias of time shift estimator $E(\hat{\epsilon})$ vs ϵ

additive white Gaussian noise, filtered by a 1-kHz low-pass filter and sampled at a rate of 2048 samples per second in an analog-to-digital converter for processing by the digital computer. A sample size of 512 was used in the computer, which began sampling at a time $T[(1/2) + \epsilon]$ relative to the start of the signal f_1 . The computer then calculated the discrete Fourier components r_1 and r_2 and from them calculated $\hat{\epsilon}$. An empirical probability distribution and the sample mean and variance of $\hat{\epsilon}$ were calculated after a large number of repetitions of the process. The sample variance obtained by experiment is compared with the predicted variance in Fig. 13. The experimental results agree closely with the predicted values.

4. Closed-Loop Synchronization

The relatively high variance of the estimate $\hat{\epsilon}$ for a single trial estimation leads to the conclusion that the estimate should be averaged over several trials in order to improve its accuracy. On the other hand, a simple averaging process performed by taking some number of independent estimates ($L > 1$) and averaging them together reduces the variance by a factor of $1/L$ but does not correct the bias of the estimate. To eliminate the bias, a synchronizing scheme should be aimed at rapidly making ϵ as close to zero as possible. To combine these objectives, the closed loop system illustrated in Fig. 16 was devised.

In this synchronization scheme, it is assumed that the pair of frequencies f_1 followed by f_2 is repeated L times before the data are transmitted. During this synchronization interval, the estimator forms L estimates, $\hat{\epsilon}_0, \hat{\epsilon}_1, \dots, \hat{\epsilon}_{L-1}$, of the instantaneous time displacement. After each estimate, a weighted average of the previous estimates is used to correct the present value of ϵ (the true starting time of the summation for the discrete Fourier transforms).

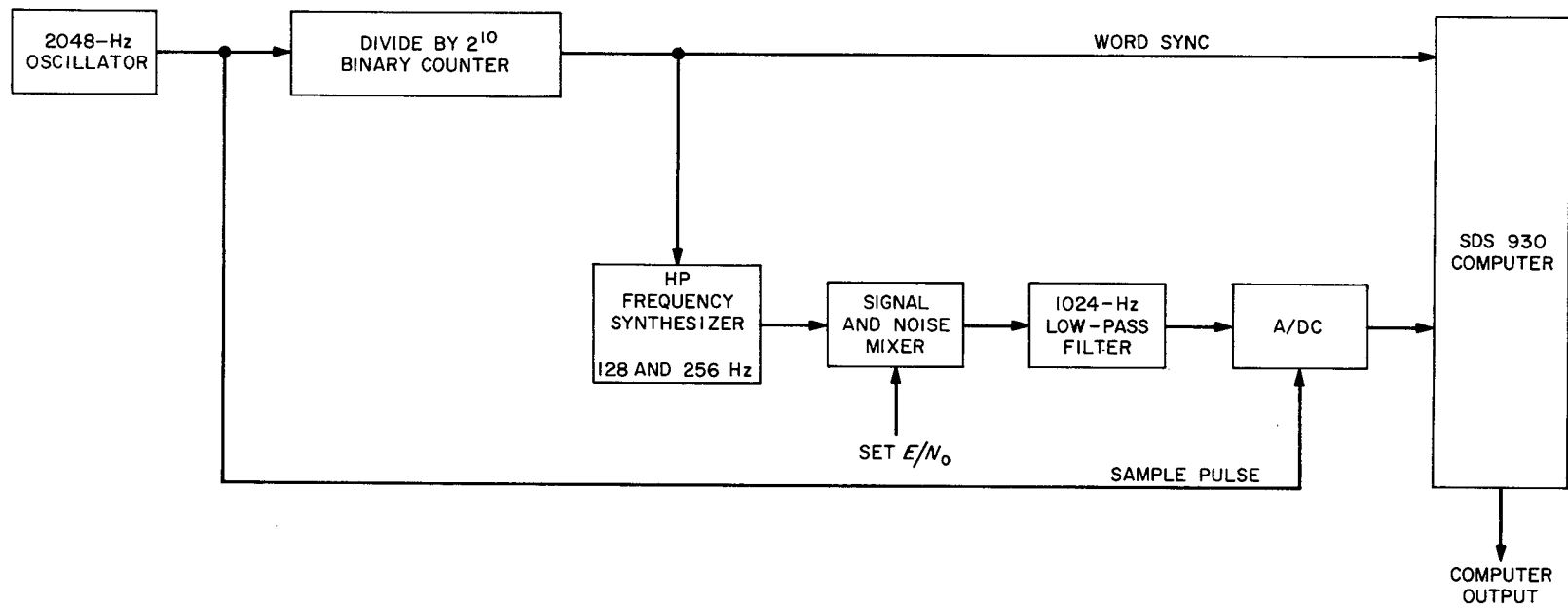


Fig. 15. Experimental setup for measurement of $P(\hat{e})$

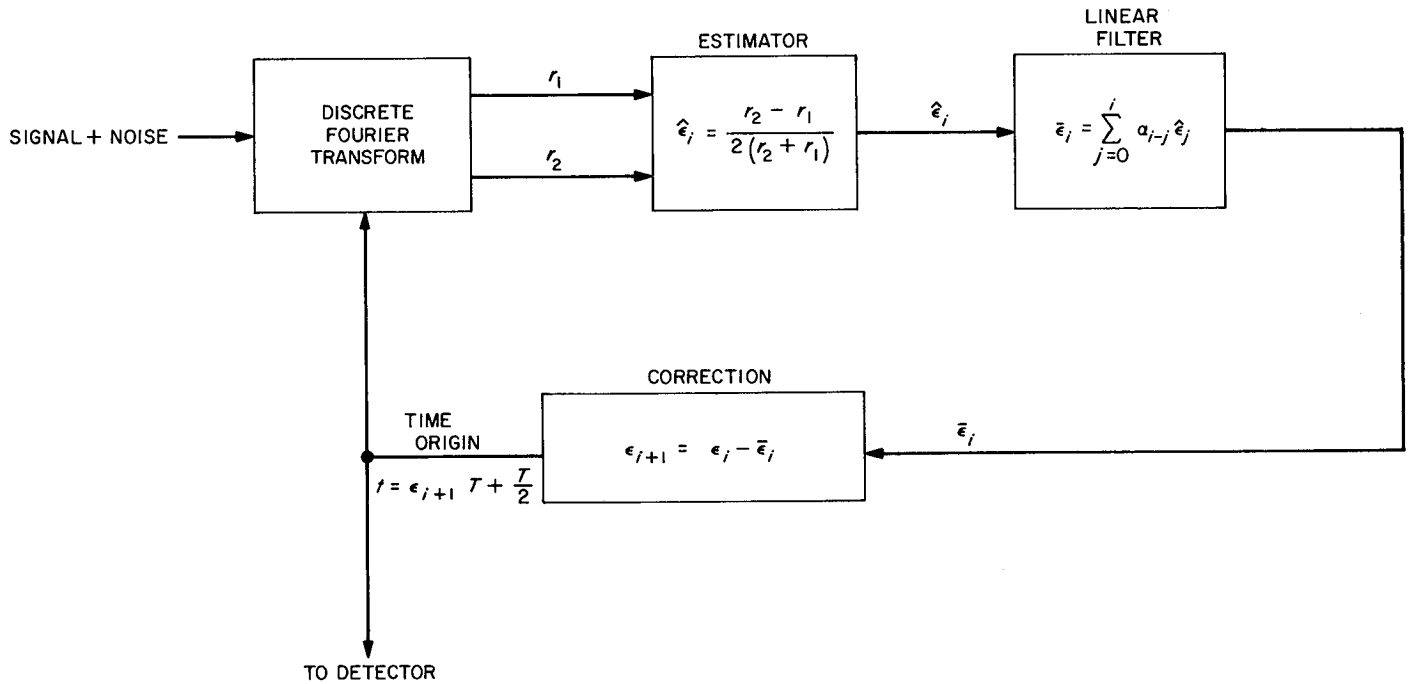


Fig. 16. Block diagram of closed-loop synchronization technique

Thus the process is a discrete-time closed-loop servo system driven toward the point $\epsilon = 0$ in L steps. Equations may be written for the value of ϵ at each step i :

$$\left. \begin{aligned} \epsilon_1 &= \epsilon_0 - a_0 \hat{\epsilon}_0 \\ \epsilon_2 &= \epsilon_1 - a_0 \hat{\epsilon}_1 - a_1 \hat{\epsilon}_0 \\ &= \epsilon_0 - (a_0 + a_1) \hat{\epsilon}_0 - a_0 \hat{\epsilon}_1 \\ &\vdots \\ \epsilon_L &= \epsilon_{L-1} - a_0 \hat{\epsilon}_{L-1} - a_1 \hat{\epsilon}_{L-2} \cdots a_{L-1} \hat{\epsilon}_0 \\ &= \epsilon_0 - (a_0 + a_1 \cdots a_{L-1}) \hat{\epsilon}_0 \\ &\quad - (a_0 + a_1 \cdots a_{L-2}) \hat{\epsilon}_1 \cdots a_0 \hat{\epsilon}_{L-1} \end{aligned} \right\} \quad (43)$$

Using this type of synchronization system as a model, it remains to determine the optimum values of the L constants a_0, a_1, \dots, a_{L-1} .

The most obvious criteria for the optimization of the system are the following:

$$\left. \begin{aligned} E(\epsilon_L) &= 0 \\ \text{var } \epsilon_L &= \text{minimum} \end{aligned} \right\} \quad (44)$$

No method for performing this optimization has been found which includes the dependence of the distribution $p(\hat{\epsilon})$ on the value of ϵ . However, by making some simplifying assumptions, it is possible to find a solution (not necessarily optimum) which does satisfy the conditions. With this solution it is then possible to determine experimentally the distribution of the final time displacement ϵ_L and, by averaging the probability of error curves over this distribution, to determine the average probability of error for a detector using this synchronization scheme. While this method is not claimed to be absolutely optimum, it is a technique that does work and leads to definite results.

a. Determination of a_0, a_1, \dots, a_{L-1} . The simplifying assumptions described above yield the following description of the estimation process:

$$\hat{\epsilon}_i = \epsilon_i + n_i \quad (45)$$

where n_i is a random variable with the characteristics⁶

⁶Under these assumptions, the estimate is unbiased and the variance is constant, independent of ϵ . A close approximation to these conditions may be made if the signal-to-noise ratio is known, by correcting the estimate $\hat{\epsilon}$ by the bias factor given by the curve of $E(\hat{\epsilon})$ vs ϵ .

$$\left. \begin{aligned} E(n_i) &= 0 \\ E(n_i n_j) &= \sigma_e^2 \delta_{ij} \end{aligned} \right\} \quad (46)$$

The solution for the values of the constants which satisfy the conditions

$$\begin{aligned} E(\epsilon_L) &= 0 \\ \text{var } \epsilon_L &= \text{minimum} \end{aligned}$$

is then

$$\left. \begin{aligned} a_0 &= \frac{1}{L} \\ a_k &= \frac{(L+1)^{k-1}}{L^{k+1}} \quad 1 \leq k \leq L-1 \end{aligned} \right\} \quad (47)$$

This statement has not been proved analytically for $L > 3$ but has been experimentally verified by the computer. The variance of the final value ϵ_L for this simplified model is reduced by a factor of $1/L$ times the variance of a single estimate. While this is an optimistic figure, as is shown by the experimental results, it is still an indication of the advantage gained by averaging the estimates by this closed-loop procedure.

b. Experimental results. A computer program has been written which simulates the action of the closed-loop synchronization scheme. The program is similar to that used to determine the density function $p(\hat{\epsilon})$, except that L successive estimations are performed and the weight average of the estimates is used to correct the value of the time displacement ϵ , as described in the preceding section. Tests were run for various signal-to-noise ratios and for different values of L , to determine the empirical probability density of the final estimate under different conditions. The initial displacement ϵ_0 was picked randomly from a uniform distribution before each trial, and 1,000 trials were run for each value of the parameters.

c. Probability of error. The curves shown in Figs. 17 and 18 are the result of numerically integrating the probability of error versus time shift curves (Figs. 10 and 11) weighted by the empirical distributions of the time shift described in the preceding section. The curves were calculated for several values of the averaging length L and for M , the number of words in the signal set equal to 2 and 32. The curve marked "perfect synchronization" is

that obtained by solving the probability of error expression given by Arthurs and Dym (Eq. 1).

5. Conclusions

A time synchronization system for an MFSK receiver has been proposed and analyzed. The analysis has been based on several simplifying assumptions, the most important of which are

- (1) The received signals are sinusoidal and known exactly in frequency and in signal length T . This assumption also means that the truncation techniques that have been proposed for the spectral analysis have not been included.
- (2) A synchronizing signal (consisting of a pair of frequencies, f_1 followed by f_2 , each lasting T seconds, and repeated several times) is transmitted before the data. (Synchronization from the data itself has not been attempted.) It is then assumed that the transmitter and receiver timing sources are stable enough that the timing estimate initially obtained will be valid for the remainder of the transmission.
- (3) In deriving the closed-loop synchronizing system, the variance of the estimate $\hat{\epsilon}$ has been assumed constant with ϵ , and the bias has been ignored.

The effect of these assumptions is, in general, to reduce the error probability of the idealized system over that of a real system. For this reason, the probability-of-error figures reached in this report may be regarded as a lower bound on the true probability of error, rather than as exact figures.

References

1. Charles, F., and Springett, J., "The Statistical Properties of the Spectral Estimates Used in the Decision Process by a Spectrum Analyzer Receiver," presented at the National Telemetry Conference, San Francisco, Calif., May 1967.
2. Arthurs, E., and Dym, H., "On the Optimum Detection of Digital Signals in the Presence of White Gaussian Noise," *IRE Trans. Comm. Sys.*, Vol. CS-10, pp. 336-372, Dec. 1962.
3. Cochran, W., et al., "What is the Fast Fourier Transform?" *IEEE Trans. Audio Electroacoust.*, Vol. AU-14, pp. 45-55, June 1967.
4. Jolley, L., *Summation of Series*, Dover Publications, Inc., New York, 1961.
5. Marcum, J., *Table of Q Functions*, Report M-339. The Rand Corporation, Santa Monica, Calif., Jan. 1, 1950.

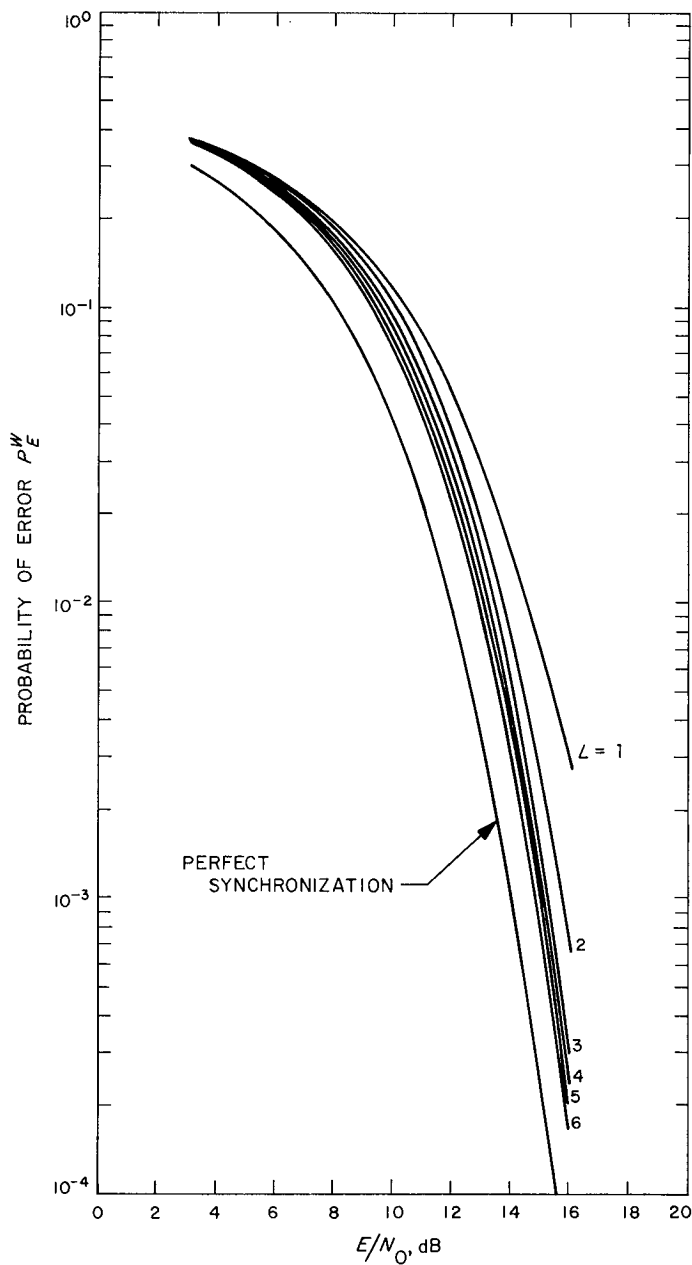


Fig. 17. Probability of error per word vs signal-to-noise ratio for closed-loop synchronizer technique. Initial timing error uniformly distributed; $m = 2$

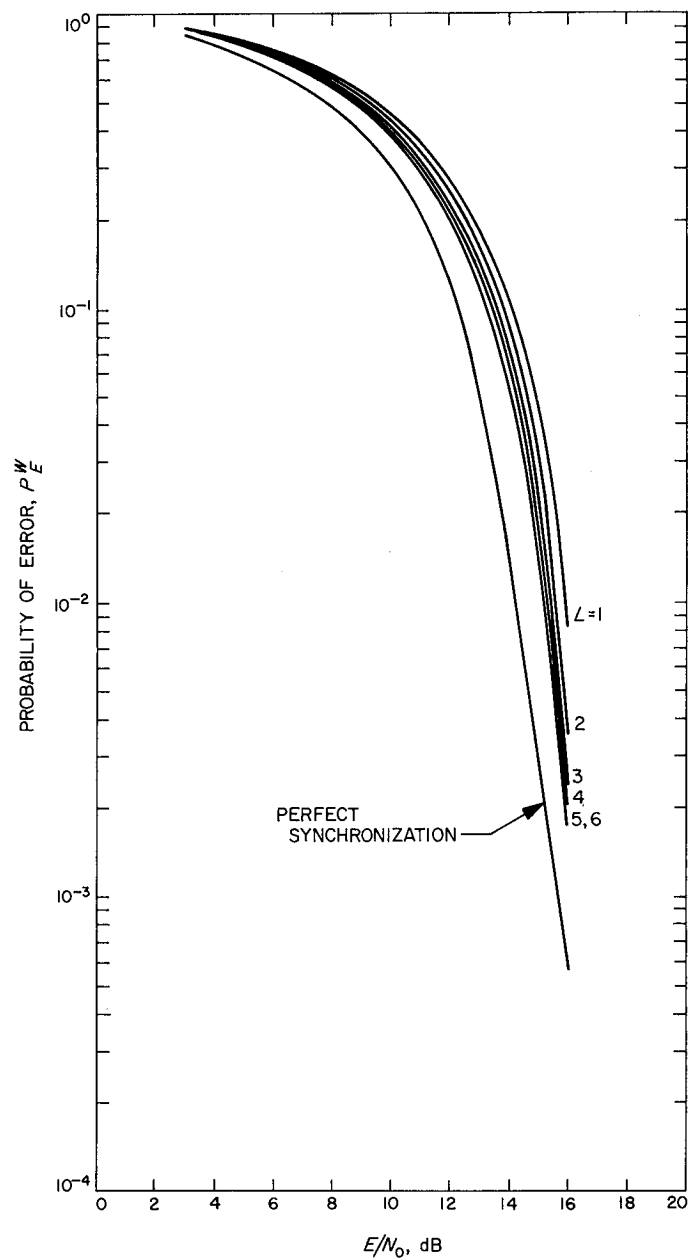


Fig. 18. Probability of error per word vs signal-to-noise ratio for closed-loop synchronizer technique. Initial timing error uniformly distributed; $m = 32$

XXIII. Spacecraft Radio

TELECOMMUNICATIONS DIVISION

A. High Impact S-Band Isolator Magnetic Materials Study, A. W. Kermode

A magnetic materials study contract (951565) was let with Rantec, Calabasas, California, a Division of Emerson Electric Company. The objective of the study was to evaluate the effects of high impact shock (10,000 *g*) and heat sterilization (135°C) on magnetic materials typically used in S-band isolators. The magnetic materials included in the study were: YIG, permanent magnets, and magnetic shielding materials.

A high impact stripline circulator was designed and utilized as a test vehicle for evaluation of circulator performance, using the various types of YIG materials.

YIG material study. Two types of S-band YIG materials from each of two manufacturers were selected for evaluation in the study program. A minimum of two samples of each material were ground to the configurations required for use in determining the basic material properties. The material parameters used for characterization of YIG are: saturation moment ($4\pi M_s$), ferrimagnetic linewidth (ΔH), dielectric constant (ϵ), gyromagnetic ratio (G_{eff}), and dielectric loss tangent ($\tan \delta$). The material

parameters are calculated using data obtained from microwave measurements. The measurements are made at X-band with an appropriate sample of YIG material inserted into an appropriate resonant cavity.

Contractor and vendor capability did not exist for measurement of linewidth or dielectric constant vs temperature. Consequently, these material parameters were only measured at room temperature.

Spherical YIG samples, 0.097 in. in diameter, were used to determine the material saturation moment and linewidth (Refs. 1 and 2). The spherical samples were high impact tested with little difficulty. The measurement results have been summarized in Table 1. The maximum change in saturation moment due to high impact was 1.9%. The change in linewidth was 27% for one sample, with the remaining samples changing less than 10%. The results show that magnetostriction did result due to high impact.

The ferrimagnetic linewidth (ΔH) is defined as the separation of the two internal static magnetic fields at which the RF power absorbed by the ferrimagnet is equal to one-half the maximum absorption. YIG materials used

Table 1. Comparison of linewidth and saturation moment measurements before and after 10,000-g impact

Material No.	Identification	-3 dB ΔH^a		-15 dB ΔH^a	G_{eff}^b	$4\pi M_s^c$
		Before	After			
MCL-1116FH	A - 1	61	62	268	2.00	570
				261	2.00	580
MCL-1116FH	A - 2	55	70	241	2.00	560
				266	2.00	570
MCL-601-5	B - 1	88	94	371	2.01	551
				423	2.01	560
MCL-601-5	B - 2	80	86	335	2.01	541
				352	2.01	551
Trans-Tech G-600	C - 1	59	61	249	2.00	657
				285	2.00	667
Trans-Tech G-600	C - 2	59	65	259	2.00	676
				286	2.00	676
Trans-Tech G-610	D - 1	44	45	203	2.00	647
				244	2.00	647
Trans-Tech G-610	D - 2	46	46	213	2.00	654
				235	2.00	654

^a ΔH = linewidth in oersteds
^b G_{eff} = Landé spectroscopic splitting factor
^c $4\pi M_s$ = saturation moment in gauss

in isolators are magnetically biased so as to be sufficiently removed from ferrimagnetic absorption, in order to assure low RF loss performance in the forward direction. The resulting magnetostriction and change in linewidth are not critical in terms of YIG material requirements for S-band isolators.

Cylindrical YIG rods, 0.050 in. in diameter and 0.650 in. in length, were used to determine the material dielectric constant (Ref. 3). Several attempts were made to high-impact the rods, resulting in the rods being fractured each time.

Changes in the dielectric constant of the YIG material used in an isolator will primarily affect the isolation characteristic. Postimpact performance results obtained on a preliminary circulator indicated that further effort to high impact YIG rods was not warranted.

Permanent magnet study. The five types of permanent magnets included in the study were: Alnico 5, Alnico 8, Alnico 9, barium ferrite, and platinum-cobalt. Two discs of each type of magnet material were ground to 0.622 in. in diameter and 0.092 in. thick. The magnets were charged magnetically and inserted as pairs into a test fixture, which duplicated the magnetic circuit used in the test circulator. The test fixture had a machined slot for insertion of a

gauss meter probe between the parallel-centered magnet stack. A separate test fixture was used to measure the magnetic field characteristics of each type of magnet pair. Measurements were made over the type approval temperature range of -10° to $+75^\circ\text{C}$, after each of six separate 6-h heat soaks at 148°C , before and after high impact tests conducted in each of two mutually perpendicular axes.

The Alnico 5 magnets could not be stabilized at the low field value (560 Gauss) and were eliminated from the test program. The results of the magnet evaluation have been summarized in Table 2. The maximum overall change in magnetic field due to all environments was 10.5%, exhibited by the Alnico 8. However, the Alnico 8 magnets exhibited the most stable field characteristics over the type approval temperature range.

Information currently available indicates that a 10.5% change in magnetic field characteristics will result in an

Table 2. Summary of permanent magnet field measurements

Test conditions	Alnico 8, G	Alnico 9, G	Nonoriented barium ferrite, G	Platinum-cobalt, G
Room ambient	570	560	580	560
$+75^\circ\text{C}$	565	540	510	550
-10°C	570	580	640	600
Room ambient	580	570	585	570
Room ambient after first 148°C cycle for 6 h	580	570	585	570
After second 148°C cycle	570	565	575	565
After fifth 148°C cycle	580	560	580	565
4 days after above measurement	620	560	580	560
Just before impact (2 days after above measurement)	625	555	580	561
After 10,000-g impact, 0.5 ms duration in two axes	628	542	581	562
7 days after impact	630	542	578	560

isolation degradation of approximately 5 dB at room temperature. The use of an appropriate magnetic stabilization procedure during charging of the magnets will reduce the relaxation effects exhibited by Alnico 8 to less than 3%, resulting in improved isolation stability.

S-band circulator. A circulator structure was used to evaluate isolator performance, using the various YIG materials. A single YIG disc was used in the circulator-design to facilitate meeting the high impact requirements. The YIG disc was centered between a pair of magnets, which were in turn centered in a magnetic yoke assembly. The yoke assembly provided a closed magnetic path, internal to the circulator.

The results of the YIG material evaluation in the test circulator have been summarized in Table 3.

Magnetic shielding study. Magnetic shielding straps were added externally to the circulator to evaluate the effective reduction and stability of the external radial magnetic field. The types of shielding straps evaluated were: 0.014-in. and 0.050-in. molypermalloy, 0.014-in. and 0.050-in. mumetal. The results of the shielding evaluation have been summarized in Table 4.

Environmental test results. The best combination of materials, resulting from the materials study, were assembled into a prototype high impact circulator structure

Table 3. Summary of circulator performance with various YIG materials

Condition	Material			
	MCL 1116 FH	MCL 601-5	TT-G 600	TT-G 610
Doping	Al	Gd-Al	Gd-Al	Al
Insertion loss, dB ^a				
Room ambient	0.30	0.27	0.30	0.20
-10°C	0.37	0.27	0.30	0.25
+75°C	0.40	0.30	0.35	0.32
Isolation, dB				
Room ambient ^b	21.7	22.0	29.0	28.2
-10°C	19.5	21.1	23.5	22.8
+75°C	19.3	19.0	25.0	18.5
^a Measurement band = 2295 ± 50 MHz				
^b Measurement band = 2295 ± 200 MHz				

Table 4. Summary of magnetic shielding measurements

Shield type	Maximum radial field at 1.5 ft, gamma				
	As received	40-G deperm	25-G exposure	80-G deperm	Maximum variation
None	18.7	14.2	6.9	12.1	11.8
0.014-in. Moly perm	2.5	2.9	1.8	1.4	1.5
0.050-in. Moly perm	2.7	3.2	2.4	2.8	0.8
0.014-in. Mumetal	1.2	2.8	1.3	1.2	1.6
0.050-in. Mumetal	2.6	3.3	2.5	2.7	0.8

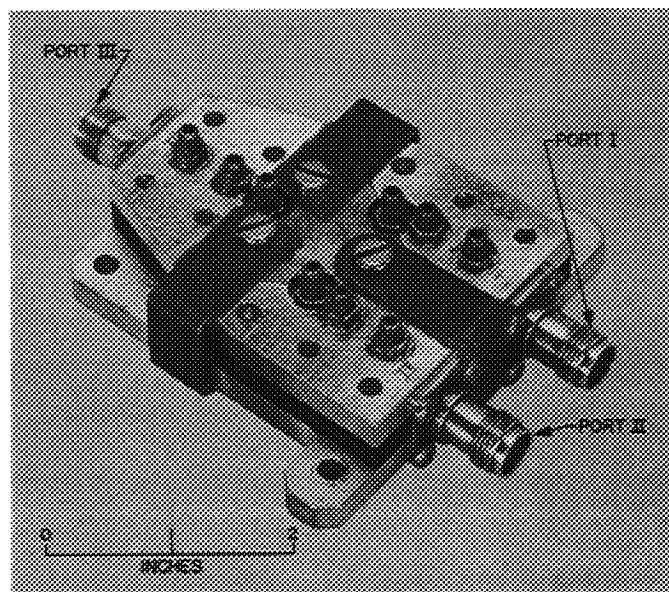


Fig. 1. Prototype high impact S-band circulator

(Fig. 1). The materials used were: Trans-Tech G-600 YIG disc, Alnico 8 permanent magnets, and 0.050-in. mumetal shielding straps. The prototype circulator was evaluated at JPL. The evaluation involved: sterilization tests of three 26-h cycles of ethylene oxide gas treatment, three 64-h cycles of 135°C heat treatment, and high-impact (10,000 g) shock tests. Magnetic mapping and electrical performance tests were made before and after the sterilization and shock tests. The insertion loss and isolation performance characteristics of the prototype circulator have been summarized in Tables 5 and 6, respectively.

During the second high impact test a failure of the mounting screws occurred, resulting in damage of two

Table 5. Summary of prototype circulator insertion loss data

Frequency, GHz	Loss, dB		
	Initial bench test	Post-ETO and heat sterilization (3 cycles each)	Post-10,000-g shock (single axis)
Ports I-II			
2.1	0.33	0.34	0.35
2.3	0.32	0.33	0.33
2.5	0.27	0.33	0.35
Ports II-III			
2.1	0.37	0.37	0.37
2.3	0.31	0.31	0.31
2.5	0.29	0.32	0.32
Ports III-I			
2.1	0.32	0.43	0.47
2.3	0.33	0.33	0.37
2.5	0.30	0.33	0.35

Table 6. Summary of prototype circulator isolation data

Frequency, GHz	Isolation, dB		
	Initial bench test	Post-ETO and heat sterilization (3 cycles each)	Post-10,000-g shock (single axis)
Ports II-I			
2.1	35.0	36.5	42.5
2.3	33.5	37.0	34.3
2.5	32.5	29.4	30.8
Ports III-II			
2.1	31.0	37.0	34.0
2.3	32.5	24.2	23.8
2.5	32.7	26.6	26.3
Ports I-III			
2.1	27.3	28.9	30.7
2.3	34.7	34.2	34.4
2.5	36.6	37.8	34.2

RF connectors. The circulator is being refurbished so that the evaluation tests can be completed.

Results to date indicate that magnetic materials are presently available that can be used in an S-band isolator capable of surviving high impact and sterilization environments.

References

1. *Test for Saturation Magnetization*, No. 663, Tech-Briefs Nos. 661-666, Test and Measurement of Ferrimagnetic Materials, Trans-Tech, Inc., Gaithersburg, Md., 1967.
2. *Test for Linewidth and Gyromagnetic Ratio*, No. 662, Tech-Briefs Nos. 661-666, Test and Measurement of Ferrimagnetic Materials, Trans-Tech, Inc., Gaithersburg, Md., 1967.
3. *Test for Complex Dielectric Constant*, No. 661, Tech-Briefs Nos. 661-666, Test and Measurement of Ferrimagnetic Materials, Trans-Tech, Inc., Gaithersburg, Md., 1967.

B. Effect of Interference on a Binary Communication Channel Using Known Signals, M. A. Koerner

1. Introduction

Many communication systems are the aggregate of one or more communication channels multiplexed to operate over the same radio link. The receivers for these communication channels are usually designed to extract information from a signal observed in white gaussian noise. In such systems, interfering signals may seriously degrade the performance of these receivers. In some cases, the interfering signal may be generated within the communication system itself. The distortion signals generated in frequency-multiplexed, PM communication systems are of this type. In other cases, the interfering signal may be generated by a second communication system operating on an adjacent frequency band. The problem common to both cases is one of evaluating the effect of the interfering signal on the performance of a receiver.

This report examines the effect of sinusoidal or gaussian interfering signals on the probability of error for a maximum-likelihood receiver for extracting binary data from a sequence of messages in white gaussian noise when each signal has duration T and is chosen randomly with equal a priori probability from a dictionary of two messages. The report initially presents equations for the form of the receiver and the probability of error for the receiver when no interfering signal is present. The remainder of the article evaluates the effect of sinusoidal and gaussian interference on the probability of error for the receiver.

2. Maximum Likelihood Receiver

Figure 2 shows a block diagram of the maximum-likelihood receiver for extracting binary data from a sequence of signals in white gaussian noise when each signal has duration T and is chosen randomly with equal a priori probability from a dictionary of two signals. If

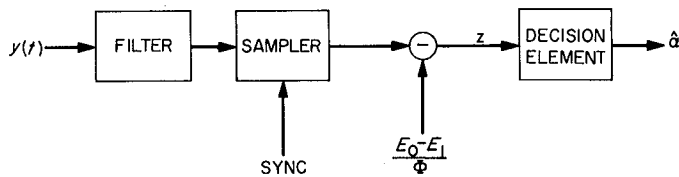


Fig. 2. Maximum-likelihood receiver functional block diagram

$s(0; t)$ and $s(1; t)$ are the two signals which can be received and Φ is the one-sided power spectral density of the white gaussian noise, the filter F has impulse response.

$$h_F(\tau) = \begin{cases} \frac{2}{\Phi} [s(0; t) - s(1; t)], & 0 \leq \tau \leq T \\ 0, & \tau > T \end{cases} \quad (1)$$

At the end of each received signal, the output of the filter F is sampled and a bias of $(E_0 - E_1)/\Phi$ is removed.

$$E_\alpha = \int_0^T s^2(\alpha; t) dt, \quad \alpha = 0, 1, \quad (2)$$

is the received signal energy. A decision element determines whether the resulting statistic z is positive or negative and sets \hat{a} , the maximum-likelihood receiver output to zero or one. If $z > 0$, $\hat{a} = 0$ and if $z < 0$, $\hat{a} = 1$.

When no interfering signal is present, the bit error probability for this receiver is

$$P_E = p(\lambda) = \frac{1}{2} [1 - \text{Erf}(\lambda^{1/2})] \quad (3)$$

where

$$\text{Erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt \quad (4)$$

and

$$\lambda = \frac{E_0 + E_1 - 2\rho(E_0 E_1)^{1/2}}{4\Phi} \quad (5)$$

The parameter ρ in Eq. (5) is the crosscorrelation between $s(0; t)$ and $s(1; t)$. In Fig. 3, $\log p(\lambda)$ is plotted as a function of $10 \log \lambda$.

3. Receiver Error Probability as a Function of the Interference-to-Signal Ratio

When either a sinusoidal or a gaussian interfering signal is present in addition to the white gaussian receiver noise,

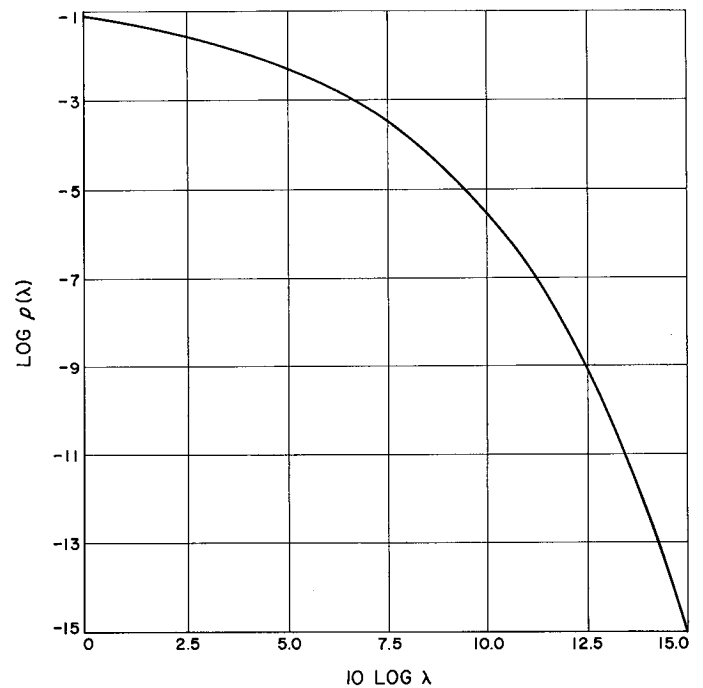


Fig. 3. $\log p(\lambda)$ as a function of $10 \log \lambda$

the receiver performance will be degraded. When a sinusoidal interfering signal having power P_i and angular frequency ω_i is present, the bit error probability for the receiver is

$$P_E = p_s(\lambda; \eta) = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1}{2} [1 - \text{Erf} \{ \lambda^{1/2} [1 + (2\eta)^{1/2} \sin u] \}] du \quad (6)$$

where, if $A_F(\omega)$ is the amplitude response of the filter F , the interference-to-signal ratio at the input to the decision element is

$$\eta = \frac{P_i A_F^2(\omega_i)}{(4\lambda)^2} \quad (7)$$

The function $\log p_s(\lambda; \eta)$ is plotted as a function of $10 \log \lambda$ for selected values of $10 \log \eta$ in Fig. 4 and as a function of $10 \log \eta$ for selected values of $10 \log \lambda$ in Fig. 5.

When a gaussian interfering signal, having two-sided power spectral density $G_i(f)$, is present

$$P_E = p_G(\lambda; \eta) = \frac{1}{2} \{1 - \text{Erf} [\lambda^{1/2} (1 + 2\eta\lambda)^{-1/2}]\} \quad (8)$$

where

$$\eta = (4\lambda)^{-2} \int_{-\infty}^{\infty} G_i(f) A_p^2(\omega) df \quad (9)$$

The function $\log p_G(\lambda; \eta)$ is plotted as a function of $10 \log \lambda$ for selected values of $10 \log \eta$ in Fig. 6 and as a function $10 \log \eta$ for selected values of $10 \log \lambda$ in Fig. 7.

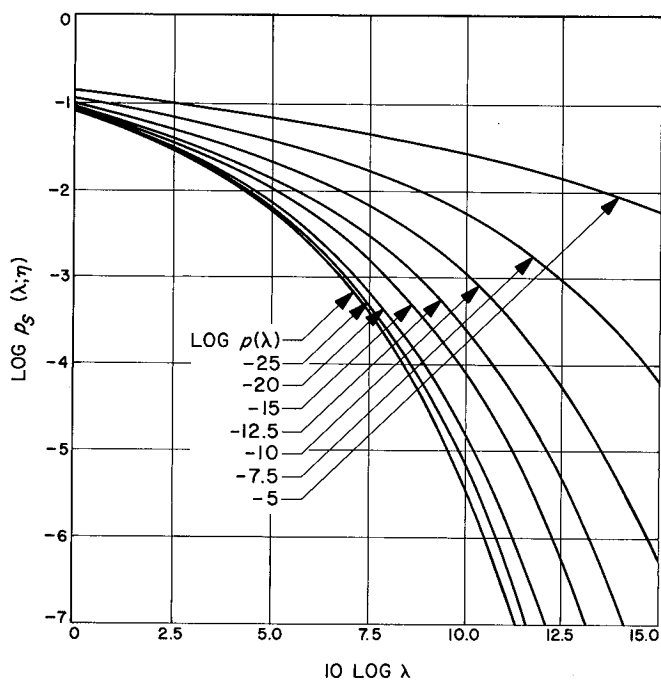


Fig. 4. $\log p_s(\lambda; \eta)$ as a function of $10 \log \lambda$ for selected values of $10 \log \eta$

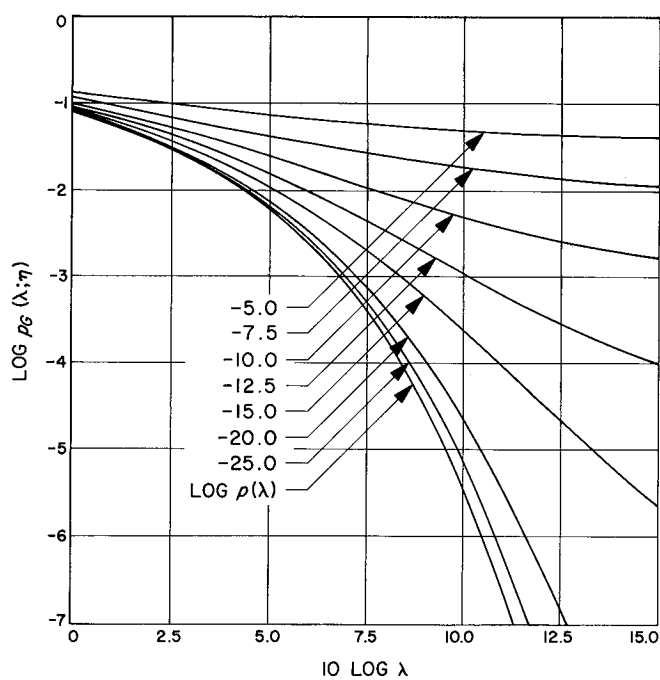


Fig. 6. $\log p_G(\lambda; \eta)$ as a function of $10 \log \lambda$ for selected values of $10 \log \eta$

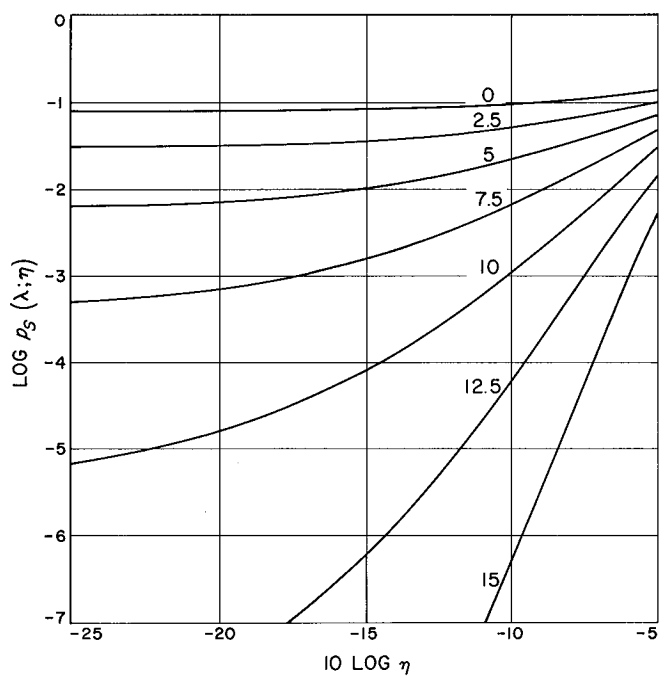


Fig. 5. $\log p_s(\lambda; \eta)$ as a function of $10 \log \pi$ for selected values of $10 \log \lambda$

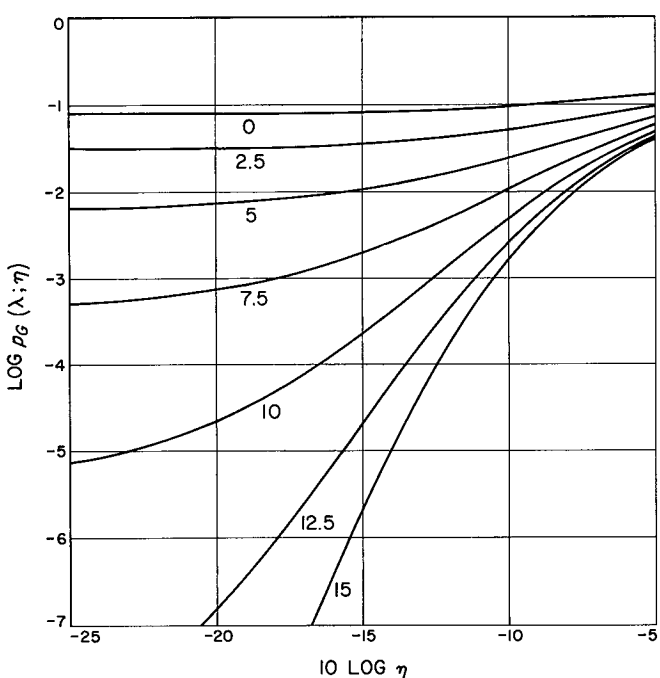


Fig. 7. $\log p_G(\lambda; \eta)$ as a function of $10 \log \eta$ for selected values of $10 \log \lambda$

4. Interference-to-Signal and Interference-to-Noise Ratios

In evaluating the effect of an interfering signal on the performance of this receiver, one finds that change in receiver bit error probability is an inconvenient measure of the receiver degradation caused by the interfering signal. Hence, we shall introduce the parameter δ , the factor by which λ must be increased to make the receiver bit error probability, when an interfering signal is present, equal to what it would be were the interfering signal absent. In most cases, δ will be a more convenient measure of receiver degradation than the change in receiver bit error probability.

Using δ as a measure of receiver degradation has the disadvantage that the value of δ depends not only on the initial values of λ and η , but also on the relationship between η and λ , as the latter parameter is increased to compensate for the presence of the interfering signal. To illustrate this point, let us examine the special case where $s(0; t)$ and $s(1; t)$ are antipodal, binary-valued signals. In this case

$$s(\alpha; t) = (-1)^\alpha P_s^{1/2} \quad (10)$$

$$E_0 = E_1 = P_s T \quad (11)$$

where P_s is the received signal power, and

$$\lambda = \frac{P_i T}{\Phi} \quad (12)$$

Then, for a sinusoidal interfering signal

$$\eta = \frac{P_i}{P_s} \frac{\sin^2(\pi f_i T)}{(\pi f_i T)^2} \quad (13)$$

while, for a gaussian interfering signal

$$\eta = \frac{P_i}{P_s} \int_{-\infty}^{\infty} \frac{G_i(f)}{P_i} \frac{\sin^2(\pi f T)}{(\pi f T)^2} df \quad (14)$$

In Fig. 8 the function $10 \log \frac{\sin^2(\pi f_i T)}{(\pi f_i T)^2}$ is plotted as a function of $f_i T$.

Examining Eqs. (12) through (14), one notes that η may either remain constant or vary as λ is increased, depending on the source of the interfering signal and which param-

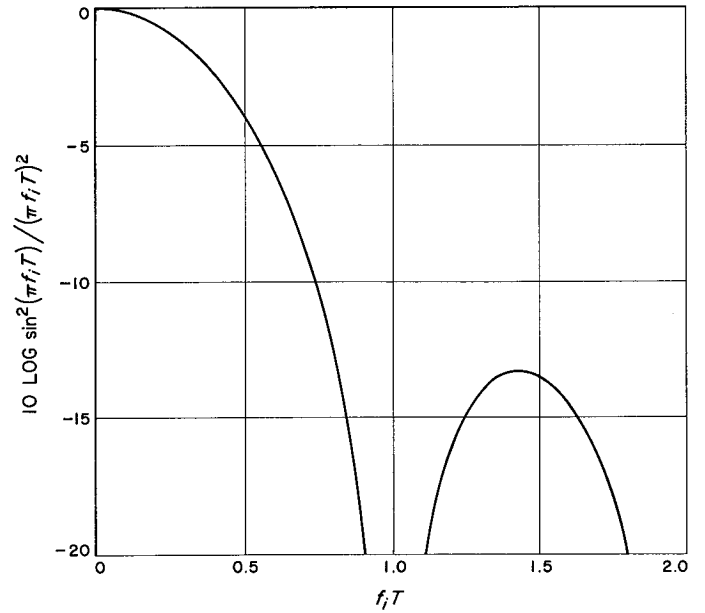


Fig. 8. $10 \log \sin^2(\pi f_i T)/(\pi f_i T)^2$ as a function of $f_i T$

eters of the communication system are changed to compensate for the degradation produced by the interfering signal. In frequency-multiplexed PM communication systems, interfering signals are generated in the process of phase modulating an RF carrier. In this case the ratio of P_i to P_s is fixed and η will remain constant. When signals are received from the transmitters for two communication systems operating on adjacent frequency bands, a portion of the signal from one transmitter may fall into the frequency band used by the other communication system. If one compensates for the degradation caused by this interfering signal by changing the receiving system parameters of the communication system, P_i/P_s and, therefore η , will remain constant. However, if one compensates for the degradations caused by this interfering signal by changing the transmitting system parameters of this communication system, P_i/P_s and therefore η will decrease as λ is increased. In the latter case the parameter remaining constant is ξ , the interference-to-noise ratio at the input of the decision element. In a communication system using antipodal, binary-valued signals,

$$\xi = 2 \frac{P_i T}{\Phi} \frac{\sin^2(\pi f_i T)}{(\pi f_i T)^2} \quad (15)$$

for sinusoidal interfering signals, while for gaussian interfering signals

$$\xi = 2 \frac{P_i T}{\Phi} \int_{-\infty}^{\infty} \frac{G_i(f)}{P_i} \frac{\sin^2(\pi f T)}{(\pi f T)^2} df \quad (16)$$

Hence, in evaluating δ , we must consider both the case where η remains constant and the case where ξ remains constant.

For arbitrary signal waveforms

$$\xi = \frac{P_i A_p^2(\omega_i)}{8\lambda} \quad (17)$$

for sinusoidal interfering signals, and

$$\xi = \frac{P_i}{8\lambda} \int_{-\infty}^{\infty} \frac{G_i(f)}{P_i} A_p^2(\omega) df \quad (18)$$

for gaussian interfering signals. Examining Eqs. (7), (9), (17) and (18), as well as Eqs. (13) through (16), one notes that

$$\xi = 2\lambda\eta \quad (19)$$

5. Receiver Error Probability as a Function of the Interference-to-Noise Ratio

Expressing the receiver bit error probability as a function of λ and ξ , for sinusoidal interfering signals the bit error probability is

$$\begin{aligned} P_E &= p_s \left(\lambda; \frac{\xi}{2\lambda} \right) \\ &= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1}{2} [1 - \text{Erf}(\lambda^{1/2} + \xi^{1/2} \sin u)] du \end{aligned} \quad (20)$$

The function $\log p_s [\lambda; \xi/(2\lambda)]$ is plotted as a function of $10 \log \lambda$ for selected values of $10 \log \xi$, in Fig. 9 and as a function of $10 \log \xi$ for selected values of $10 \log \lambda$ in Fig. 10.

For gaussian interfering signals

$$\begin{aligned} p_E &= p_g \left(\lambda; \frac{\xi}{2\lambda} \right) \\ &= \frac{1}{2} \{1 - \text{Erf}[\lambda^{1/2} (1 + \xi)^{-1/2}]\} \end{aligned} \quad (21)$$

The function $\log p_g [\lambda; \xi/(2\lambda)]$ is plotted as a function $10 \log \lambda$ for selected values of $10 \log \xi$ in Fig. 11 and as a function of $10 \log \xi$ for selected values of $10 \log \lambda$ in Fig. 12.

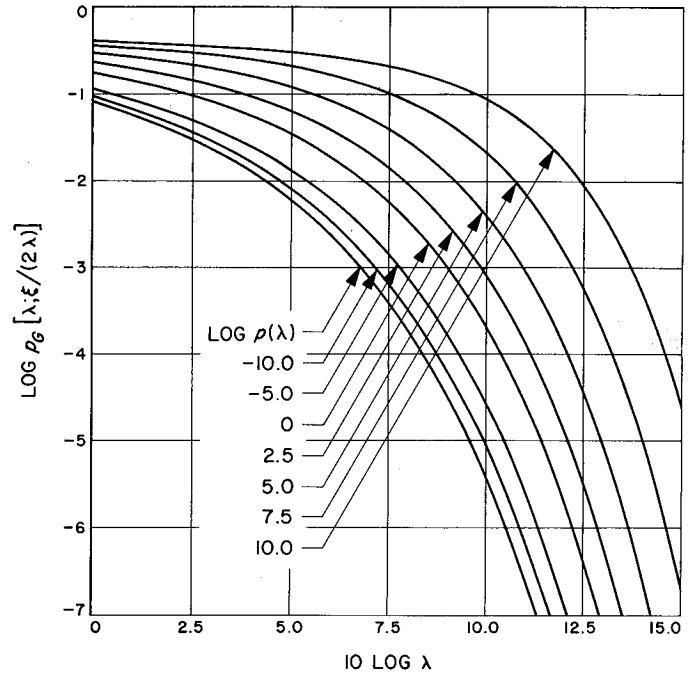


Fig. 9. $\log p_s [\lambda; \xi/(2\lambda)]$ as a function of $10 \log \lambda$ for selected values of $10 \log \xi$

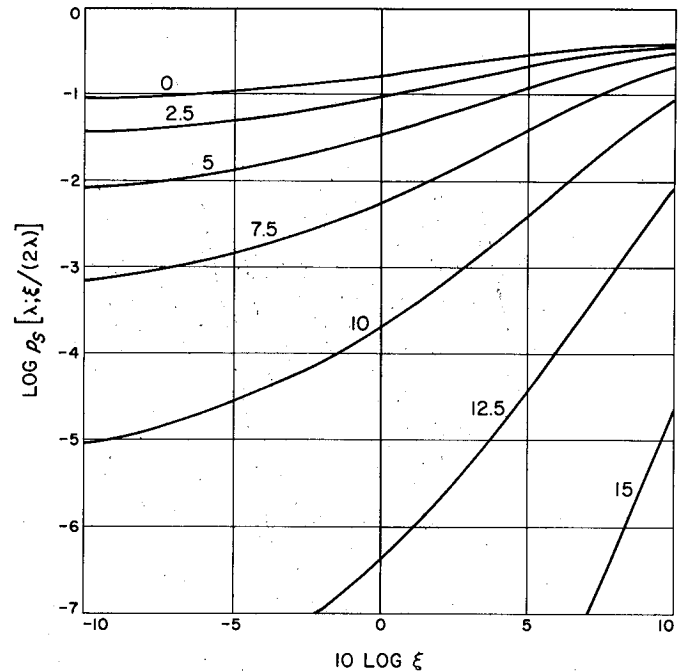


Fig. 10. $\log p_s [\lambda; \xi/(2\lambda)]$ as a function of $10 \log \xi$ for selected values of $10 \log \lambda$

6. Receiver Degradation

a. *Sinusoidal interference, constant interference-to-signal ratio.* Since the factor δ is the amount λ must be

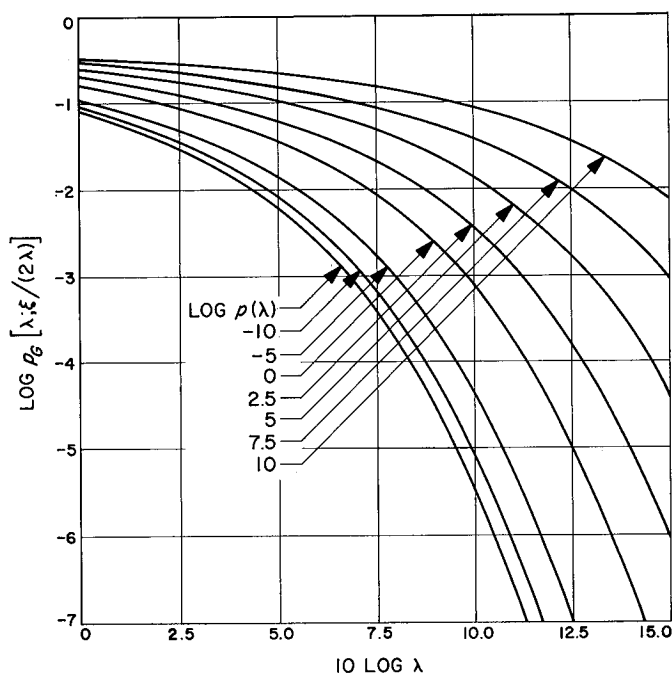


Fig. 11. $\text{Log } p_G [\lambda; \xi/(2\lambda)]$ as a function of $10 \log \lambda$ for selected values of $10 \log \xi$

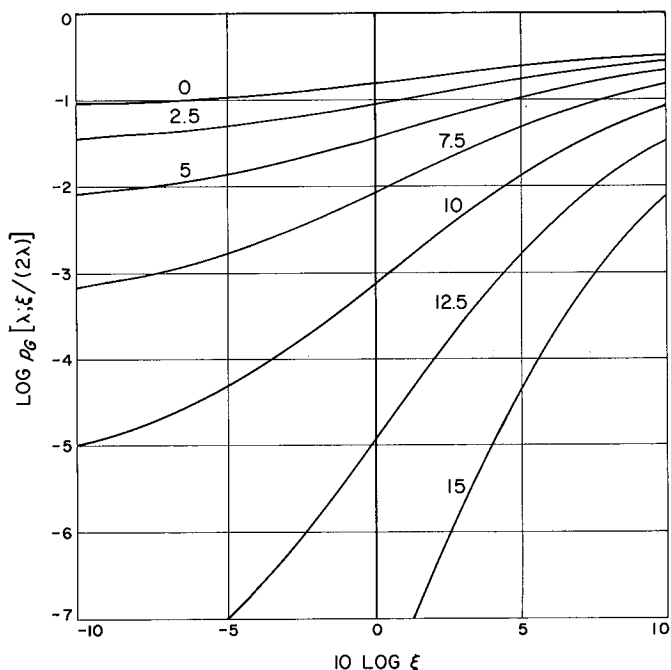


Fig. 12. $\text{Log } p_G [\lambda; \xi/(2\lambda)]$ as a function of $10 \log \xi$ for selected values of $10 \log \lambda$

increased to compensate for the presence of the interfering signal, when η is fixed and the interference is sinusoidal, δ is the solution of the equation

$$p_S(\delta\lambda; \eta) = p(\lambda) \quad (22)$$

or, using Eqs. (3) and (6),

$$\begin{aligned} \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1}{2} [1 - \text{Erf} \{ \delta^{1/2} \lambda^{1/2} [1 + (2\eta)^{1/2} \sin u] \}] du \\ = \frac{1}{2} [1 - \text{Erf} (\lambda^{1/2})] \end{aligned} \quad (23)$$

$10 \log \delta$ is plotted as a function of $10 \log \lambda$ for selected values of $10 \log \eta$ in Fig. 13 as a function of $10 \log \eta$ for selected values of $10 \log \lambda$ in Fig. 14. Since

$$\lim_{\lambda \rightarrow \infty} p_S(\lambda; \eta) = \begin{cases} 0, & \eta \leq \frac{1}{2} \\ \frac{1}{2} - \frac{1}{\pi} \sin^{-1} [(2\eta)^{-1/2}], & \eta > \frac{1}{2} \end{cases} \quad (24)$$

a finite solution of Eq. (23) for δ will exist for all values of λ when $\eta < 1/2$ and for $\lambda < \lambda_0$, where

$$\text{Erf} (\lambda_0^{1/2}) = \frac{2}{\pi} \sin^{-1} [(2\eta)^{-1/2}] \quad (25)$$

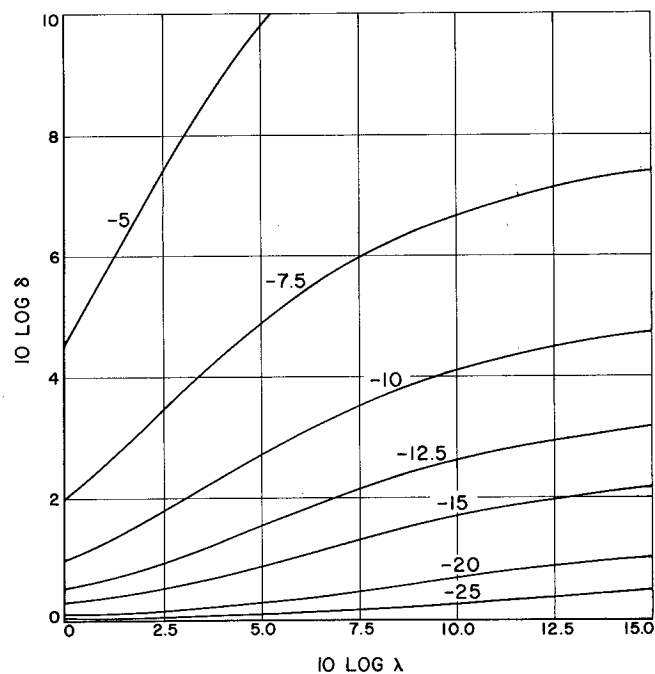


Fig. 13. $10 \log \delta$ for sinusoidal interference and constant η as a function of $10 \log \lambda$ for selected values of $10 \log \eta$

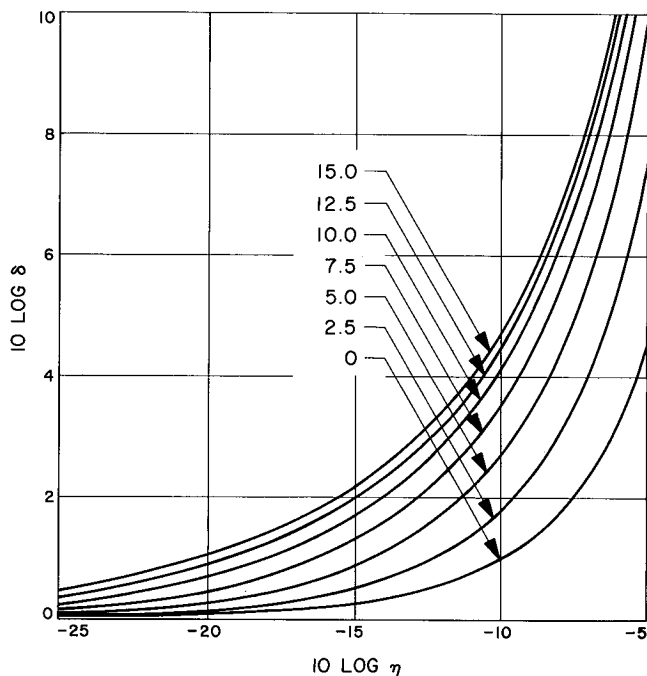


Fig. 14. $10 \log \delta$ for sinusoidal interference and constant η as a function of $10 \log \eta$ for selected values of $10 \log \lambda$

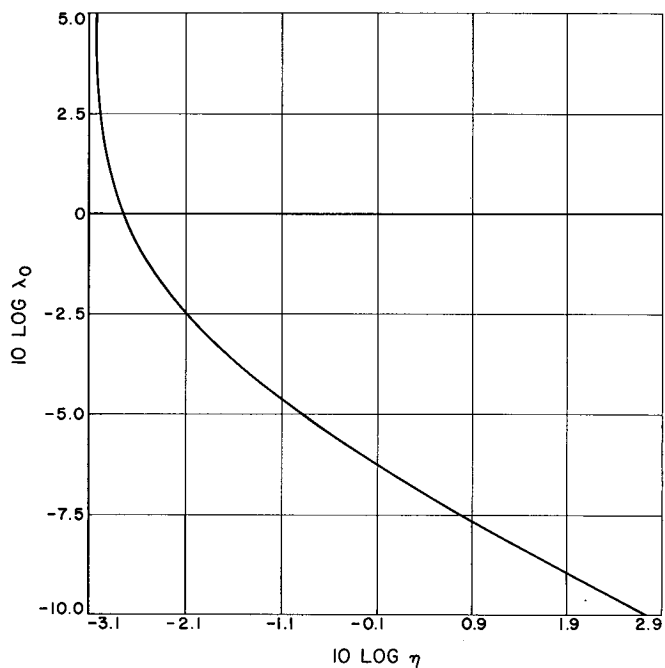


Fig. 15. $10 \log \lambda_0$ for sinusoidal interference and constant η as a function of $10 \log \eta$

when $\eta \geq 1/2$. For cases where a solution of Eq. (23) does not exist ($\lambda > \lambda_0$), δ is infinite. In Fig. 15, $10 \log \lambda_0$ is plotted as a function of $10 \log \eta$.

b. Gaussian interference, constant interference-to-signal ratio. When η is fixed, and a gaussian interfering signal is present, δ is the solution of the equation.

$$p_G(\delta\lambda; \eta) = p(\lambda) \quad (26)$$

Since

$$\lim_{\lambda \rightarrow \infty} p_G(\lambda; \eta) = \frac{1}{2} \{1 - \text{Erf}[(2\eta)^{-1/2}]\} \quad (27)$$

for $\lambda > \lambda_0 = (2\eta)^{-1}$, δ is infinite, while

$$\delta = (1 - 2\eta\lambda)^{-1}, \quad \lambda < \lambda_0 = (2\eta)^{-1} \quad (28)$$

$10 \log \delta$ is plotted as a function of $10 \log \lambda$ for selected values of $10 \log \eta$ in Fig. 16 and as a function of $10 \log \eta$ for selected values of $10 \log \lambda$ in Fig. 17.

c. Sinusoidal interference, constant interference-to-noise ratio. When ξ is fixed, for sinusoidal interfering signals δ is the solution of the equation

$$p_S\left(\delta\lambda; \frac{\xi}{2\delta\lambda}\right) = p(\lambda) \quad (29)$$

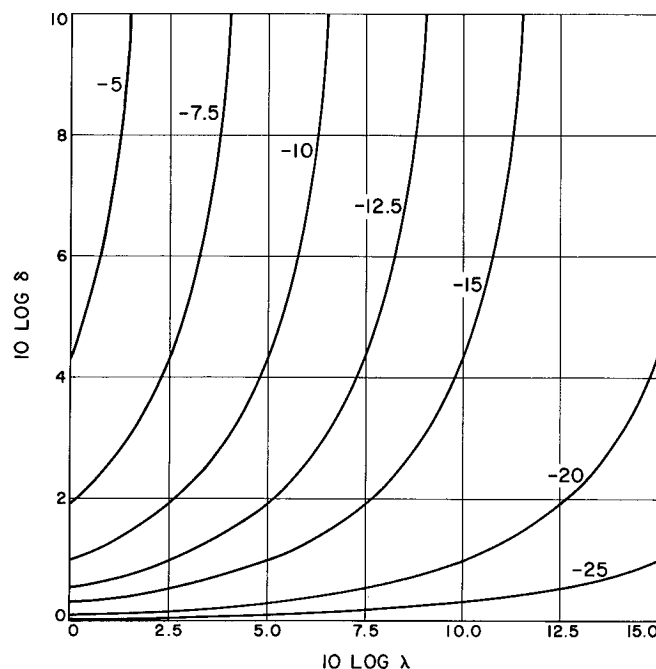


Fig. 16. $10 \log \delta$ for gaussian interference and constant η as a function of $10 \log \lambda$ for selected values of $10 \log \eta$

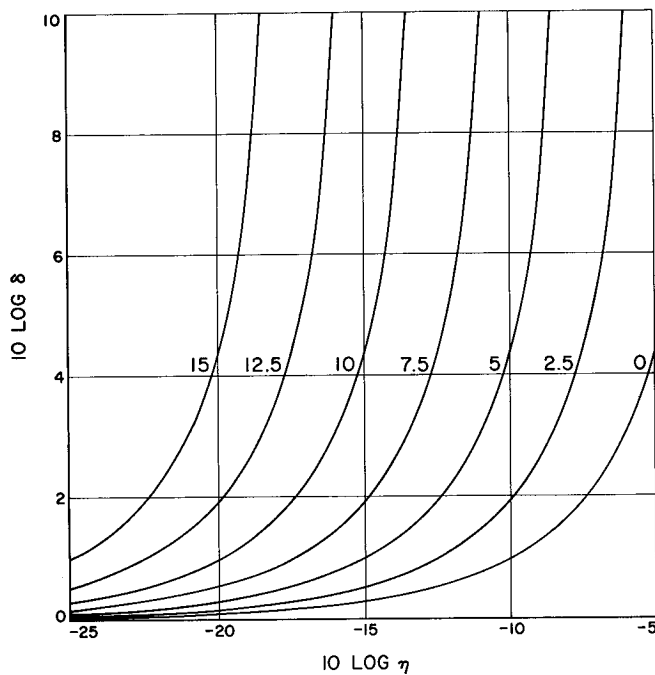


Fig. 17. $10 \log \delta$ for gaussian interference and constant η as a function of $10 \log \eta$ for selected values of $10 \log \lambda$

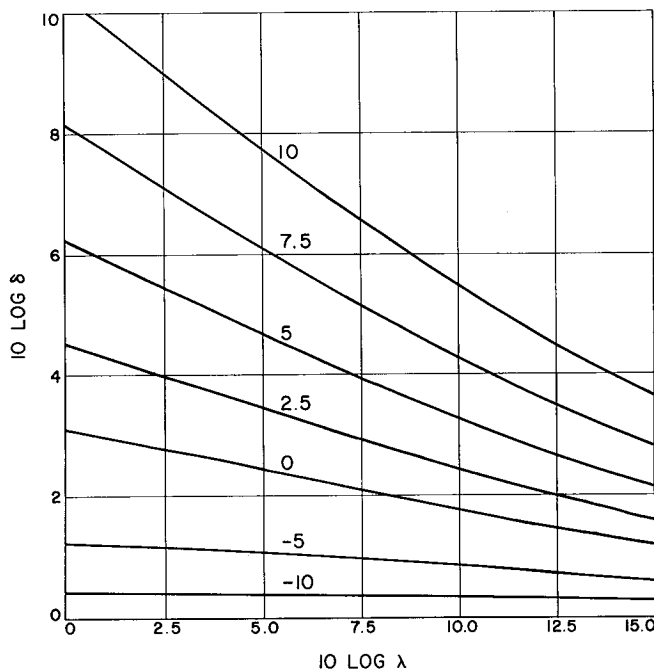


Fig. 18. $10 \log \delta$ for sinusoidal interference and constant ξ as a function of $10 \log \lambda$ for selected values of $10 \log \xi$

or, using Eqs. (3) and (20),

$$\frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1}{2} [1 - \text{Erf}(\delta^{1/2} \lambda^{1/2} + \xi^{1/2} \sin u)] du = \frac{1}{2} [1 - \text{Erf}(\lambda^{1/2})] \quad (30)$$

$10 \log \delta$ is plotted as a function of $10 \log \lambda$ for selected values of $10 \log \xi$ in Fig. 18 and as a function of $10 \log \xi$ for selected values of $10 \log \lambda$ in Fig. 19.

d. Gaussian interference, constant interference-to-noise ratio. For gaussian interfering signals, δ is the solution of the equation

$$p_g\left(\delta\lambda; \frac{\xi}{2\delta\lambda}\right) = p(\lambda) \quad (31)$$

or, using Eqs. (3) and (21),

$$\delta = 1 + \xi \quad (32)$$

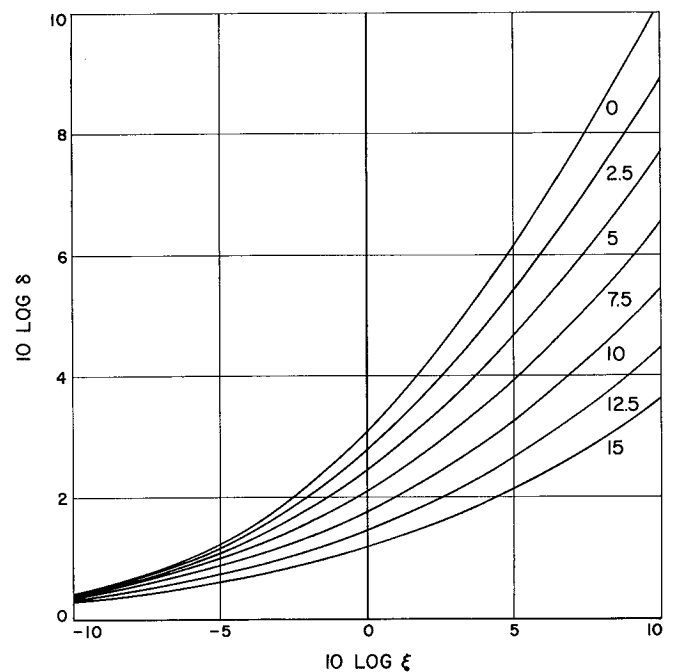


Fig. 19. $10 \log \delta$ for sinusoidal interference and constant ξ as a function of $10 \log \xi$ for selected values of $10 \log \lambda$

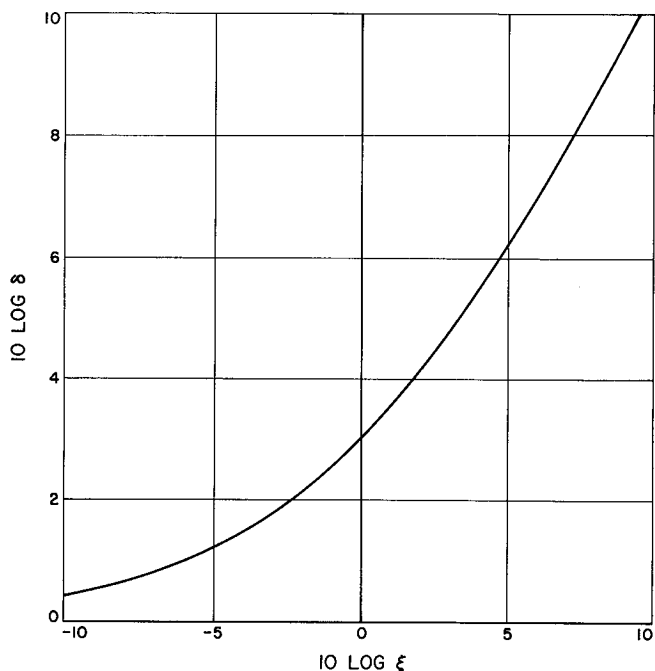


Fig. 20. $10 \log \delta$ for gaussian interference and constant ξ as a function of $10 \log \xi$

In Fig. 20, $10 \log \delta$ is plotted as a function of $10 \log \xi$. One should note that in this case δ is not dependent on λ .

Since

$$\lim_{\lambda \rightarrow \infty} p_s\left(\lambda; \frac{\xi}{2\lambda}\right) = \lim_{\lambda \rightarrow \infty} p_g\left(\lambda; \frac{\xi}{2\lambda}\right) = 0 \quad (33)$$

where ξ is fixed, δ is finite for all values of λ .

7. Comparison of the Effect of Sinusoidal and Gaussian Interference

A convenient approximation often used to evaluate the effect of a nongaussian interfering signal on the performance of a receiver is to assume that the effect of the interfering signal is the same as that of a gaussian process which produces equal power at the receiver output. In Figs. 21 through 27, we compare the behavior of the receiver error probability for sinusoidal and gaussian interference as a function of $10 \log \eta$ and $10 \log \xi$ for values $10 \log \lambda$ in the 0.0 to 15.0-dB range. The obvious conclusion is that for sinusoidal interference, the gaussian approximation is satisfactory for small λ and η or ξ but breaks down for large λ and large η or ξ .

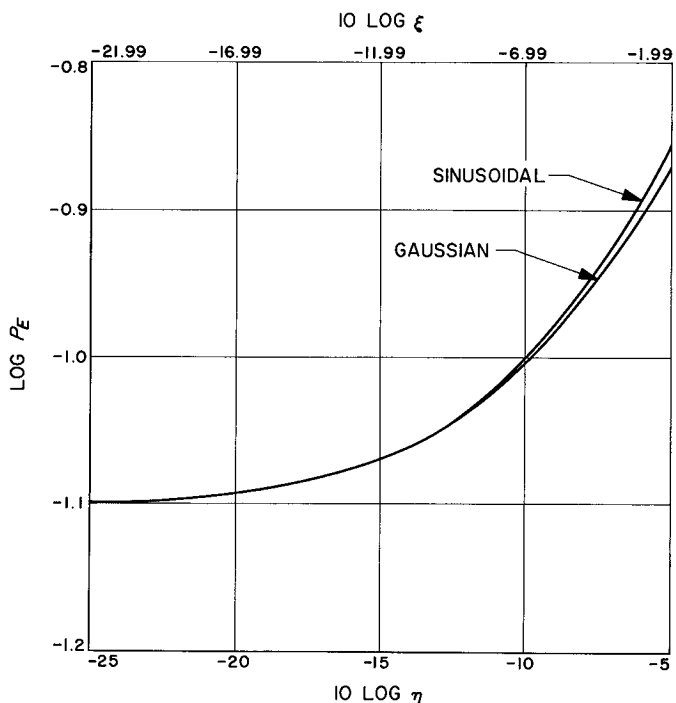


Fig. 21. Comparison of the receiver error probability for sinusoidal and gaussian interference as a function of $10 \log \eta$ and $10 \log \xi$ for $10 \log \lambda = 0.0$

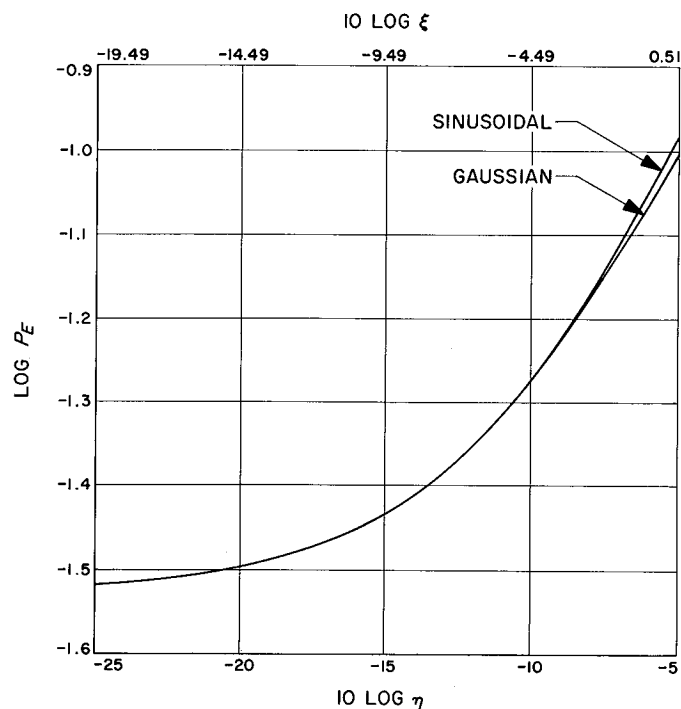


Fig. 22. Comparison of receiver error probability for sinusoidal and gaussian interference as a function of $10 \log \eta$ and $10 \log \xi$ for $10 \log \lambda = 2.5$

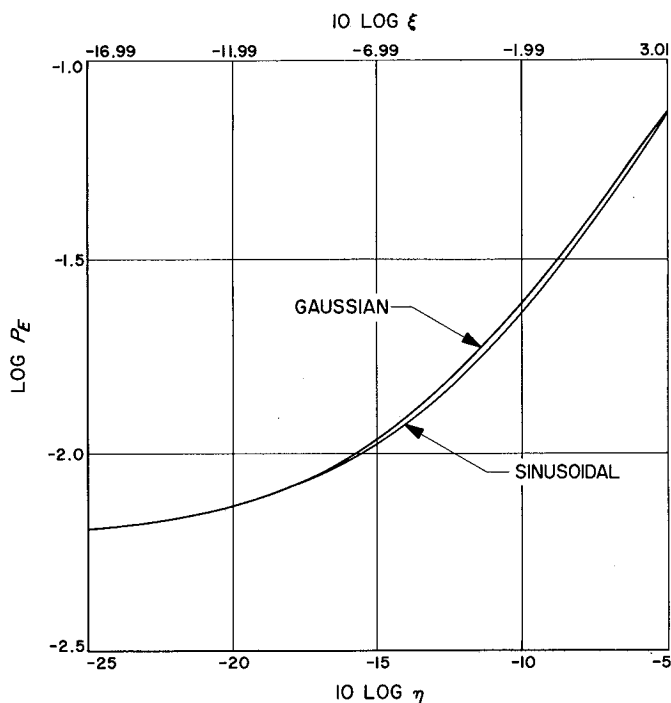


Fig. 23. Comparison of receiver error probability for sinusoidal and gaussian interference as a function of $10 \log \eta$ and $10 \log \xi$ for $10 \log \lambda = 5.0$

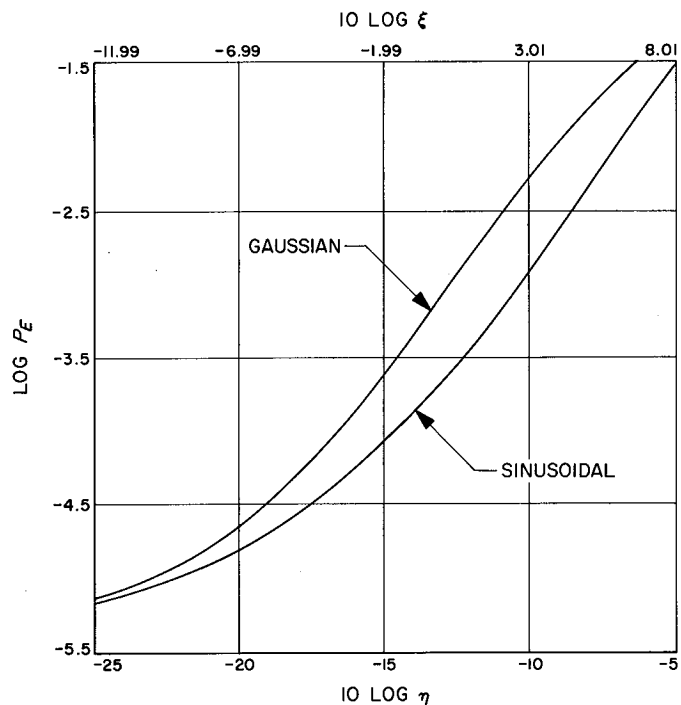


Fig. 25. Comparison of receiver error probability for sinusoidal and gaussian interference as a function of $10 \log \eta$ and $10 \log \xi$ for $10 \log \lambda = 10.0$

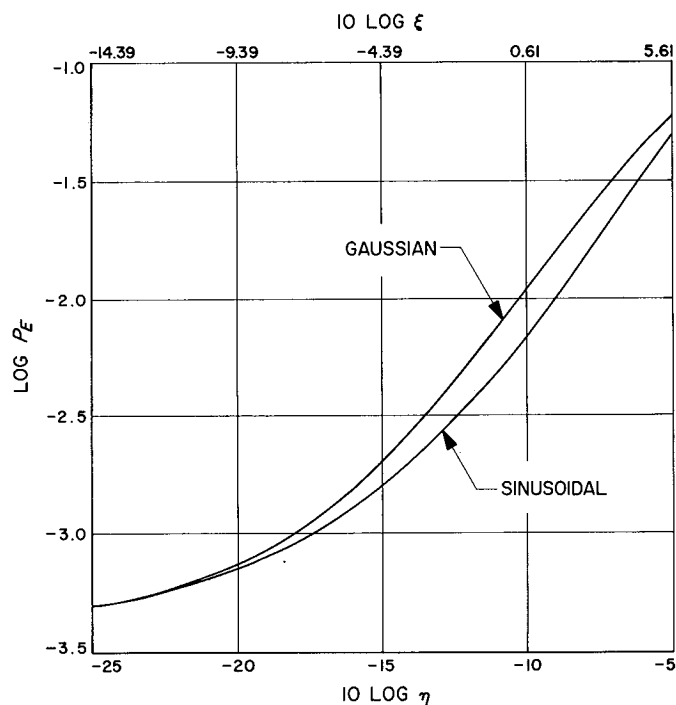


Fig. 24. Comparison of receiver error probability for sinusoidal and gaussian interference as a function of $10 \log \eta$ and $10 \log \xi$ for $10 \log \lambda = 7.5$

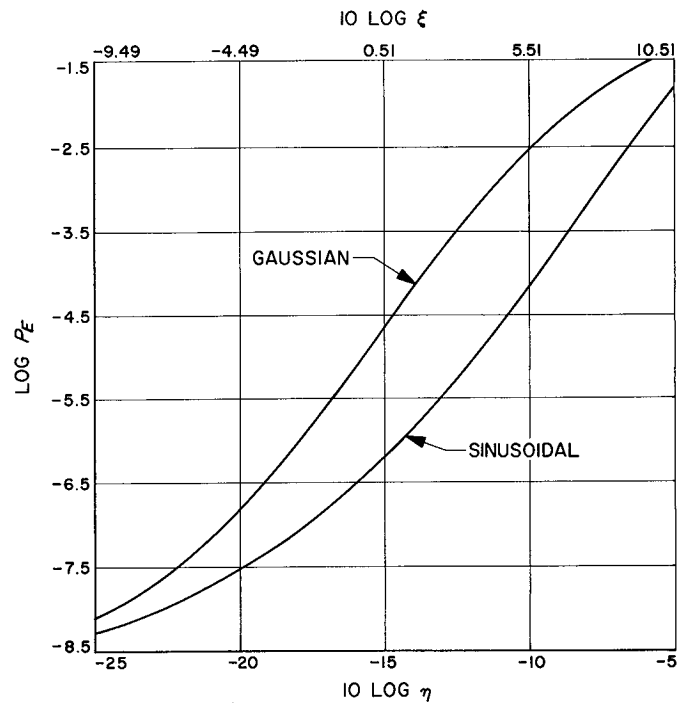


Fig. 26. Comparison of receiver error probability for sinusoidal and gaussian interference as a function of $10 \log \eta$ and $10 \log \xi$ for $10 \log \lambda = 12.5$

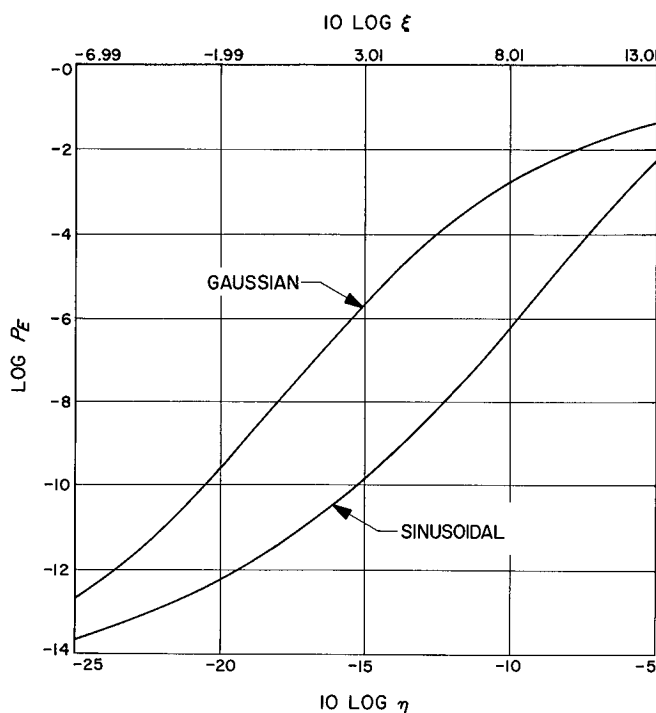


Fig. 27. Comparison of receiver error probability for sinusoidal and gaussian interference as a function of $10 \log \eta$ and $10 \log \xi$ for $10 \log \lambda = 15.0$

8. Conclusion

This article has presented the results of an analysis of the effects of sinusoidal and gaussian interference on the performance of the maximum-likelihood receiver for extracting binary data from a sequence of messages in white gaussian noise when each signal has duration T and is chosen with equal a priori probability from a dictionary of two messages.

The report presents equations for the receiver error probability and the receiver degradation as a function of the parameters λ and η or ξ . Graphs are included which show the behavior of the receiver error probability and the receiver degradation as a function of $10 \log \lambda$ and $10 \log \eta$ or $10 \log \xi$. This report also presents equations which relate the parameters λ , η , and ξ to the basic parameters of the signal, interference, and noise. Finally, a comparison is made of the effect of a sinusoidal interfering signal with that of a gaussian interfering signal. The comparison shows that for small values of λ and η or ξ , the degradation produced by sinusoidal interference is close to that produced by gaussian interference. However, the approximation is not valid for large λ or large η or ξ .

C. Spacecraft Power Amplifier Development Program, L. J. Derr

1. Electrostatically Focused Amplifier

This S-band (2295-MHz) power amplifier project is a portion of JPL's advanced development program for space-borne transmitter tubes. The work is being performed by the Klystron Department of EIMAC, division of Varian Associates, under JPL contract 951105 (SPS 37-37, Vol. IV, pp. 258-259). This article gives the current development status and summarizes the progress made.

a. Specifications. The development effort for this device started in May 1965. An overall design was chosen which would make this tube an ideal spacecraft transmitter in many ways. Electrostatic focusing was selected to avoid troublesome magnetic leakage fields for compatibility with spacecraft mounted magnetometers. Radiation cooling was specified to minimize heat contributions to the spacecraft's structure. High efficiency was specified to ease prime power requirements. Power output variability was required, a natural characteristic of electrostatically focusing klystrons, to make the tube adaptable to a wide variety of spacecraft designs. Wide bandwidth was provided to fit the modulation requirements of foreseeable future missions. RF drive requirements were kept within the range of simple, long life, solid-state circuits. This device, by all analytical criteria, should perform well for 20,000 h or more.

The important design goals for this development are repeated here for convenient reference:

- Power
output . . . 20 to 100 W (variable with beam voltage)
- Focusing . . . Electrostatic
- Efficiency . . . 35% at 20 W, 45% at 100 W
- Gain 30 dB minimum
- Bandwidth . . 30 MHz (3 dB) at all power levels
- Cooling 60% radiation, 40% conduction

b. Completed tasks. The project has, thus far, consisted of the following tasks:

- (1) Computer analysis of the electron beam dynamics in a periodic electric field (focusing)
- (2) Beam analyzer tests to determine the optimum circuit geometry for best dc to RF conversion (efficiency)
- (3) Computer analysis of staggered tuning patterns (bandwidth)

- (4) Experimental life tests of high temperature materials to be used for the radiation cooled collector (collector design)
- (5) Experimental radiating collector assemblies (radiation efficiency)
- (6) Helical circuit studies and analysis (new circuit designs)
- (7) Experimental tube assemblies (performance studies)

c. Development status. Most major design parameters have now reached the specified levels and are discussed in some detail below:

Power output. The 5 to 1 variability specification was easily met by this design. In practice, all four experimental tubes were capable of operating at any RF power level from 1 to 130 W. Only at levels below 10 W do the bandwidth and efficiency fall below the specified performance. In tubes where brazing problems caused excessive RF circuit losses power fading was noted at 130 W output, but tubes of normal construction have been power-stable at all levels of operation.

Focusing. Because of lack of previous work in this area, a large amount of theoretical analysis and empirical testing was necessary before the first experimental model could be designed and built. Although the beam was well controlled in this tube, it soon became apparent that the interaction of focusing fields, RF fields, lens aberrations, space charge forces and beam perveance was not well understood. A computer study was then implemented to determine the action of the beam under these influential factors. The study quite reliably predicted the observed performance conditions existing in the first 75% of the tube's length but failed to accurately describe the beam in the output section where high RF defocusing fields are produced. This problem was partially resolved by the beam analyzer tests.¹

The studies did not produce a classic solution to the focusing problems but were technically useful in subsequent tube designs where from 98 to 100% beam transmission was observed under dc conditions. It is intended that the final focusing lenses will be actually tied to the cathode potential, thereby requiring no power supply of their own. This configuration allows the focusing field to

change as the beam voltage is changed and thus provides the power variability feature of this design.

Efficiency. The efficiency goal has been the most difficult to achieve. In the first experimental tube a considerable portion of the beam was intercepted by the output cavity circuit. This degraded the RF beam efficiency, producing insufficient efficiency levels of only 23% at 20 W and 33% at 100 W. The output circuit tunnels were enlarged and flared in tube No. 2, which did not greatly reduce interception but did indicate an improving trend. The beam efficiencies observed were 25% at 20 W and 37% at 100 W.

Tube No. 3 was a planned experiment to resolve the beam interception problem. First a tube, using all improvements known at that time, was assembled and tested. The output cavity was then removed, and the remainder of the amplifier was placed in the vendor's beam analyzer. Here the beam was mapped under actual RF conditions, and the data used to redesign the output cavity.

Tube No. 4 incorporated the new output cavity design which resulted in a large improvement in beam efficiency. The measured values were 34% at 20 W, and 46.5% at 100 W of output power.

Gain. The gain of electrostatically focused amplifiers is consistently higher than classical klystron theory predicts. Consequently, this parameter has exceeded the specifications on all experimental tubes. Typical of this is tube No. 4 which had a gain of 35.6 dB at 20 W and 47 dB at 130 W. Its input-output characteristics, at these levels, are shown in Fig. 28. A drive power variation of > 6 dB is shown to be possible before a 1.0 dB variation in power output is observed at any level of operation.

Bandwidth. Stagger tuning of the helical resonators and the wide band characteristics of the dual output cavity have resulted in a 30-MHz bandwidth (3 dB) at all power levels from 10 to 130 W. Bandpass ripples of 0.5 dB can be seen at the lower power levels, but Q adjustments in future tubes will improve the flatness of this characteristic.

Cooling. All experimental tubes have, thus far, used water cooling to lower development costs. The radiating collector design has been tested separately, as stated in SPS 37-37, Vol. IV, pp. 258-259, and will be incorporated in the next experimental tube design. Some modifications to its entrance aperture will now be necessary. Test results on tube No. 4 show that only 60% of the electron beam reaches the inside of the collector shell when all electrode

¹These two studies are treated in detail in EIMAC's Quarterly Reports 1 and 2.

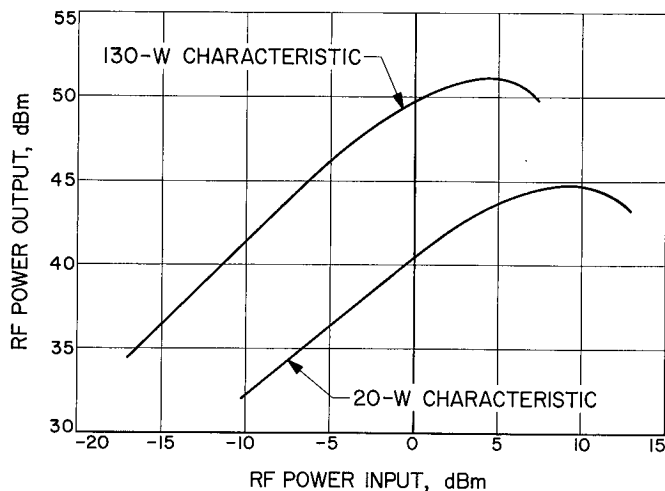


Fig. 28. Power input-output characteristics of tube No. 4, 20- to 100-W electrostatically focused amplifier

voltages are adjusted for the highest RF efficiency. The radiating collector is 80% efficient. The overall radiation cooling efficiency could, therefore, be only 48%, whereas 60% was specified.

d. Remaining tasks. Two more experimental tubes remain to be built. These units will use radiation cooling, and the collector design will be modified to improve the cooling efficiency. A final package design will be created for tube No. 6 where mounting and encapsulation problems are yet to be solved.

e. Development status. Development of the basic 20-100 W electrostatically focused amplifier is nearly complete, and most specified performance goals have been reached. This development project was originally scheduled to be completed in 18 mo, but due to the lack of experience with this configuration analytical and experimental efforts were more than anticipated. As a result, the project has been extended to 33 mo. The final tube is scheduled for delivery to JPL in March 1968.

D. Life Test Data Acquisition System, R. S. Hughes

In view of the expanding life test program in the spacecraft radio area, the need for an automatic data acquisition system became evident. This system was necessary to periodically measure and record the operating parameters of spacecraft radio components, such as RF power amplifiers on life test. This article briefly describes the data acquisition system used for this purpose, its operation, and

its capability when used in conjunction with an existing on-lab computer.

The pertinent operating parameters of the life test items can be converted to dc analog voltages, and thus a dc measuring system was adequate for this application. The prime requirements for the system were $\pm 0.01\%$ accuracy, good reliability, versatility, and the capability of correlating data and time. In addition, since the system would not contain a computer, the output data compiled by the system were to be compatible with the input to an existing on-lab computer, primarily the IBM 7094. This was necessary to facilitate the conversion of the raw data into the desired engineering units and format. Also, the utilization of the computer permits the accumulated data to be plotted at regular intervals.

This combination of requirements led to a relatively straightforward data acquisition system, as illustrated in Fig. 29. A block diagram of this data acquisition system

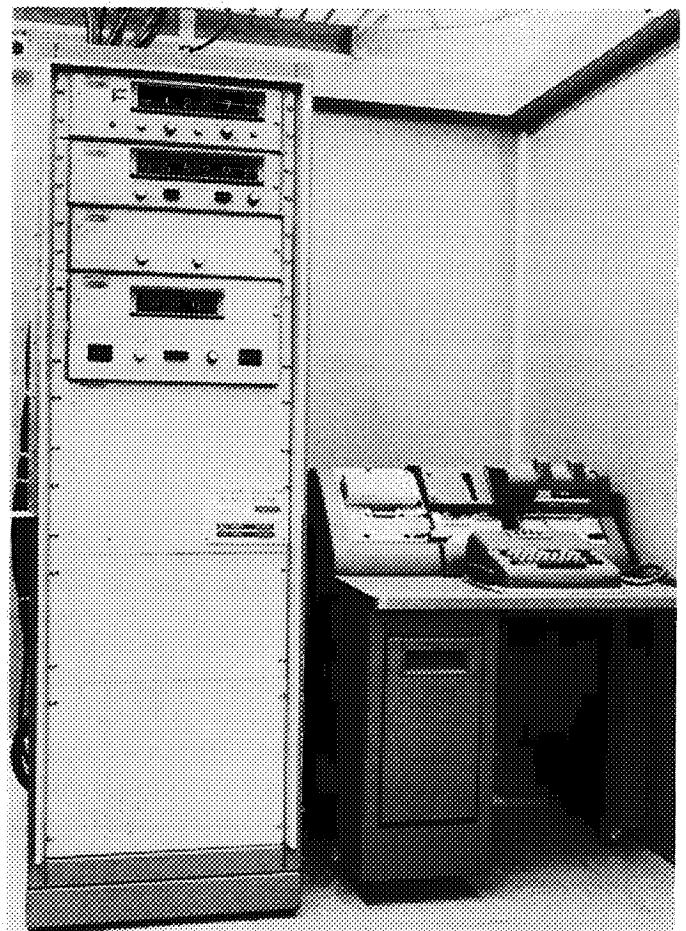


Fig. 29. Life test data acquisition system

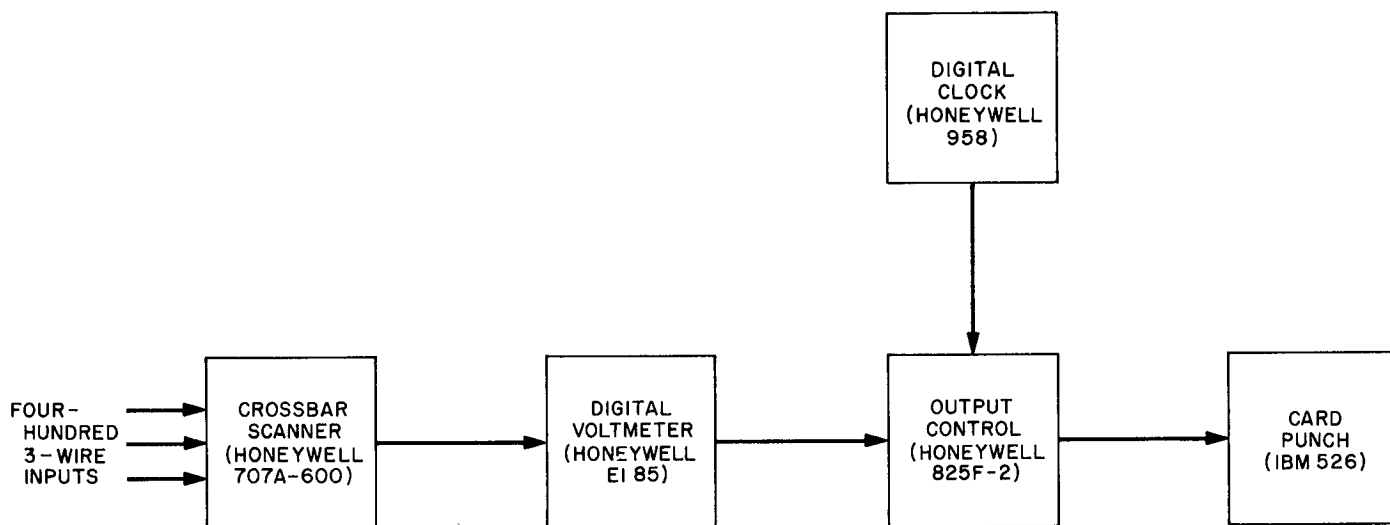


Fig. 30. Block diagram of life test data acquisition system

is shown in Fig. 30. The switching mechanism in the scanner primarily consists of two 600-point cross-bar switches which provide four-hundred 3-wire input channels. The primary function of the scanner is to sequentially select the desired channels and supply the voltages on these channels to the DVM. The ranges on the DVM extend from 100 mV to 1000 V full scale, with a typical 6-mo accuracy of $\pm 0.01\%$. The DVM automatically selects the appropriate range for the applied input signal, digitizes the signal, and supplies it to the output control. The output control converts the parallel data from the DVM, scanner, and clock to serial data and controls the operation of the card punch. The digital clock is used to correlate the data and time; for reference, clock data are placed on each card. Also, the clock commands the system to record data at prescribed intervals. In addition, the system can be commanded manually.

As a means of regularly checking the condition of the data acquisition system and its environment, a set of reference conditions are recorded at the beginning of each scan. These reference conditions include a measurement of the DVM's zero, on the most sensitive range, and several voltages derived from a zener diode reference standard. These voltages step the DVM through all ranges with the exception of the 1000-V scale. Also, a temperature measurement is taken of the system's surrounding environment.

During normal operation, the clock originates a pulse at prescribed intervals which commands the system to measure and record on IBM cards the analog voltages of

the life test items. If deemed appropriate, comment cards can be inserted between data cards to explain any anomaly, such as a power line failure. These comments will appear in the processed data, as illustrated in Fig. 31.

These compiled, raw data which may contain from one to one hundred fifty data sets are processed by the IBM 7094 computer in accordance with the computer program designed for this purpose. The computer converts the raw data into the desired engineering units and format, and it provides a conventional computer print-out. In addition, the computer punches cards containing the processed data. The program is written so that these punched cards are automatically arranged in chronological groups for each individual life test item. Therefore, the punched cards can be easily separated into groups, and the processed data for each life test item stored in individual areas. This technique facilitates the listing and plotting of the accumulated data without reprocessing the old data. In addition, this permits comment cards to be inserted in the processed data to explain any irregularities which appear in the data. These comments will then appear when the data are listed, as shown in Fig. 31.

This procedure requires a separate computer program for listing and plotting the accumulated, processed data. However, the versatility gained is well worthwhile, and in addition it conserves computer time. A computer plot displaying typical data obtained with the data acquisition system is shown in Fig. 32. Linear interpolation is used between the data points.

LIFE TEST DATA, 10 WATT TWTA, HUGHES MODEL 216H S/N 30 AND EMPS MODEL L21A S/N 16056, CHANNELS 007 TO 018																	
CLOCK HOURS	OFF HOURS	ON HOURS	PS VOLT	PS AMPS	EF VOLT	EA VOLT	EH VOLT	IH MA	EC VOLT	IC MA	PIN WATTS	PO WATTS	D PO DB	TEMP F	H PWR WATTS	C PWR WATTS	OVERALL EFF PERCENT
4489.0	127.6	22154.5	40.2	1.38	5.02	88.	1213.	6.8	855.	39.1	0.0284	10.91	-0.52	93.	8.24	33.42	25.25
4495.0	127.6	22160.5	40.2	1.38	5.02	88.	1215.	7.1	853.	39.5	0.0274	10.72	-0.60	93.	8.68	33.70	24.40
4501.0	127.6	22166.5	40.2	1.38	5.03	88.	1215.	7.4	854.	39.1	0.0280	10.64	-0.63	92.	9.04	33.37	24.19
4507.0	127.6	22172.5	40.2	1.38	5.04	88.	1214.	7.0	850.	39.4	0.0285	10.74	-0.59	92.	8.55	33.53	24.62
4513.0	127.6	22178.5	40.2	1.38	5.03	88.	1214.	7.0	851.	39.5	0.0284	10.76	-0.58	93.	8.55	33.59	24.64
4519.0	127.6	22184.5	40.2	1.38	5.03	88.	1214.	7.0	849.	39.2	0.0284	10.78	-0.57	92.	8.48	33.27	24.89
4525.0	127.6	22190.5	40.2	1.38	5.03	88.	1213.	7.0	848.	38.8	0.0282	10.69	-0.61	92.	8.48	32.89	24.89
4531.0	127.6	22196.5	40.2	1.38	5.05	88.	1213.	7.2	851.	39.5	0.0276	10.66	-0.62	92.	8.68	33.59	24.33
4536.6	127.6	22202.1	40.2	1.38	5.04	88.	1215.	7.1	853.	39.1	0.0273	10.68	-0.62	93.	8.69	33.35	24.49
4537.0	127.6	22202.5	40.2	1.38	5.03	88.	1216.	7.2	855.	39.1	0.0275	10.67	-0.62	93.	8.71	33.42	24.42
4543.0	127.6	22208.5	40.2	1.38	5.04	88.	1213.	7.1	850.	39.6	0.0279	10.72	-0.60	92.	8.68	33.62	24.45
4549.0	127.6	22214.5	40.2	1.38	5.04	88.	1214.	7.0	849.	39.2	0.0284	10.79	-0.57	92.	8.51	33.29	24.89
4555.0	127.6	22220.5	40.2	1.38	5.04	88.	1214.	6.9	850.	38.6	0.0275	10.75	-0.59	92.	8.43	32.77	25.13
THIS COMMENT CARD WAS INSERTED IN THE RAW DATA FOR ILLUSTRATION ONLY. IT APPEARS WHEN THE RAW DATA IS FIRST PROCESSED AND ALSO WHEN THE PROCESSED DATA IS LISTED.																	
4561.0	127.6	22226.5	40.2	1.38	5.04	88.	1214.	7.0	850.	38.8	0.0276	10.73	-0.59	92.	8.54	33.01	24.89
4567.0	127.6	22232.5	40.2	1.38	5.04	88.	1214.	7.0	850.	39.6	0.0280	10.69	-0.61	92.	8.52	33.69	24.43
4573.0	127.6	22238.5	40.2	1.38	5.05	88.	1213.	7.1	850.	39.7	0.0274	10.65	-0.63	91.	8.67	33.74	24.23
4579.0	127.6	22244.5	40.2	1.38	5.06	88.	1215.	6.9	850.	39.4	0.0272	10.68	-0.61	91.	8.36	33.52	24.59
4579.7	127.6	22245.2	40.2	1.38	5.05	88.	1215.	6.8	851.	39.8	0.0290	10.68	-0.62	91.	8.28	33.83	24.45
4580.6	127.6	22246.1	40.2	1.38	5.06	88.	1213.	7.0	849.	39.2	0.0282	10.66	-0.62	91.	8.54	33.26	24.58
4580.6	127.6	22246.3	40.2	1.38	5.06	88.	1213.	7.2	850.	39.7	0.0279	10.61	-0.64	91.	8.71	33.74	24.10
4580.8	127.6	22246.3	40.2	1.38	5.06	88.	1213.	7.1	851.	39.6	0.0290	10.61	-0.64	91.	8.58	33.71	24.20
4583.0	127.6	22248.5	40.2	1.38	5.04	88.	1214.	7.1	849.	38.7	0.0279	10.65	-0.63	91.	8.59	32.85	24.76
4587.0	127.6	22252.5	40.2	1.38	5.04	88.	1213.	7.2	851.	39.6	0.0280	10.61	-0.64	92.	8.74	33.67	24.13
4591.0	127.6	22256.5	40.2	1.38	5.05	88.	1211.	7.2	850.	39.3	0.0285	10.55	-0.67	91.	8.74	33.36	24.15
4595.0	127.6	22260.5	40.2	1.38	5.05	88.	1213.	7.2	850.	39.7	0.0280	10.55	-0.67	91.	8.75	33.78	23.92
4611.0	127.6	22276.5	40.2	1.38	5.00	88.	1213.	7.1	850.	39.4	0.0289	10.69	-0.61	93.	8.62	33.51	24.48
4615.0	127.6	22280.5	40.2	1.38	5.03	88.	1214.	7.1	850.	39.6	0.0277	10.58	-0.66	92.	8.59	33.67	24.14
4623.0	127.6	22288.5	40.2	1.38	5.05	88.	1214.	7.1	850.	39.7	0.0277	10.53	-0.67	90.	8.59	33.70	24.03
4627.0	127.6	22292.5	40.2	1.38	5.06	88.	1216.	7.0	855.	39.2	0.0278	10.64	-0.63	90.	8.47	33.49	24.44
4635.0	127.6	22300.5	40.2	1.38	5.04	88.	1211.	7.2	850.	39.1	0.0277	10.57	-0.66	91.	8.73	33.24	24.30
4639.0	127.6	22304.5	40.2	1.38	5.05	88.	1214.	6.9	857.	39.1	0.0278	10.64	-0.63	91.	8.35	33.53	24.50
4643.0	127.6	22308.5	40.2	1.38	5.05	88.	1215.	7.1	854.	39.2	0.0277	10.58	-0.66	90.	8.61	33.48	24.23
4651.0	127.6	22316.5	40.2	1.38	5.05	88.	1214.	7.1	851.	39.5	0.0281	10.55	-0.67	90.	8.68	33.65	24.03
4655.0	127.6	22320.5	40.2	1.38	5.05	88.	1213.	7.1	850.	39.3	0.0278	10.58	-0.66	91.	8.68	33.44	24.22
THIS COMMENT CARD WAS INSERTED IN THE PROCESSED DATA FOR ILLUSTRATION ONLY. IT APPEARS WHEN THE PROCESSED DATA IS LISTED.																	
4659.0	127.6	22324.5	40.2	1.38	5.04	88.	1213.	7.1	849.	39.6	0.0277	10.60	-0.65	91.	8.60	33.60	24.22
4663.0	127.6	22328.5	40.2	1.38	5.04	88.	1213.	7.1	850.	39.6	0.0276	10.57	-0.66	91.	8.62	33.63	24.13
4667.0	127.6	22332.5	40.2	1.38	5.05	88.	1210.	7.2	849.	39.2	0.0277	10.65	-0.63	91.	8.73	33.26	24.46
4700.1	127.6	22365.6	40.2	1.38	5.07	88.	1213.	7.2	851.	39.4	0.0282	10.56	-0.66	90.	8.74	33.58	24.06
4703.0	127.6	22368.5	40.2	1.38	5.05	88.	1215.	6.9	855.	39.1	0.0279	10.66	-0.62	91.	8.38	33.40	24.59
4709.0	127.6	22374.5	40.2	1.38	5.04	88.	1213.	7.1	850.	39.5	0.0278	10.58	-0.65	91.	8.67	33.56	24.17
4715.0	127.6	22380.5	40.2	1.38	5.04	88.	1213.	7.2	850.	39.4	0.0278	10.53	-0.68	91.	8.74	33.49	24.05
4721.0	127.6	22386.5	40.2	1.38	5.05	88.	1213.	7.2	851.	39.3	0.0280	10.53	-0.67	90.	8.73	33.42	24.10
4725.1	127.6	22390.6	40.2	1.38	5.05	88.	1215.	6.9	852.	38.8	0.0279	10.60	-0.65	91.	8.42	33.05	24.62
4728.1	127.6	22393.6	40.2	1.38	5.05	88.	1213.	7.1	849.	39.5	0.0278	10.63	-0.63	91.	8.59	33.54	24.34
4733.0	127.6	22398.5	40.2	1.38	5.04	88.	1214.	7.0	850.	39.4	0.0291	10.66	-0.62	91.	8.55	33.51	24.44
4739.0	127.6	22404.5	40.2	1.38	5.05	88.	1213.	7.2	851.	39.2	0.0279	10.56	-0.66	91.	8.73	33.37	24.19
4744.4	127.6	22409.9	40.2	1.38	5.05	88.	1213.	7.1	850.	39.7	0.0276	10.56	-0.66	90.	8.58	33.71	24.09

Fig. 31. Computer print-out of TWT life test data

LIFE TEST DATA, HUGHES TWT MODEL 216H, S/N 30 NOV. 10, 1967

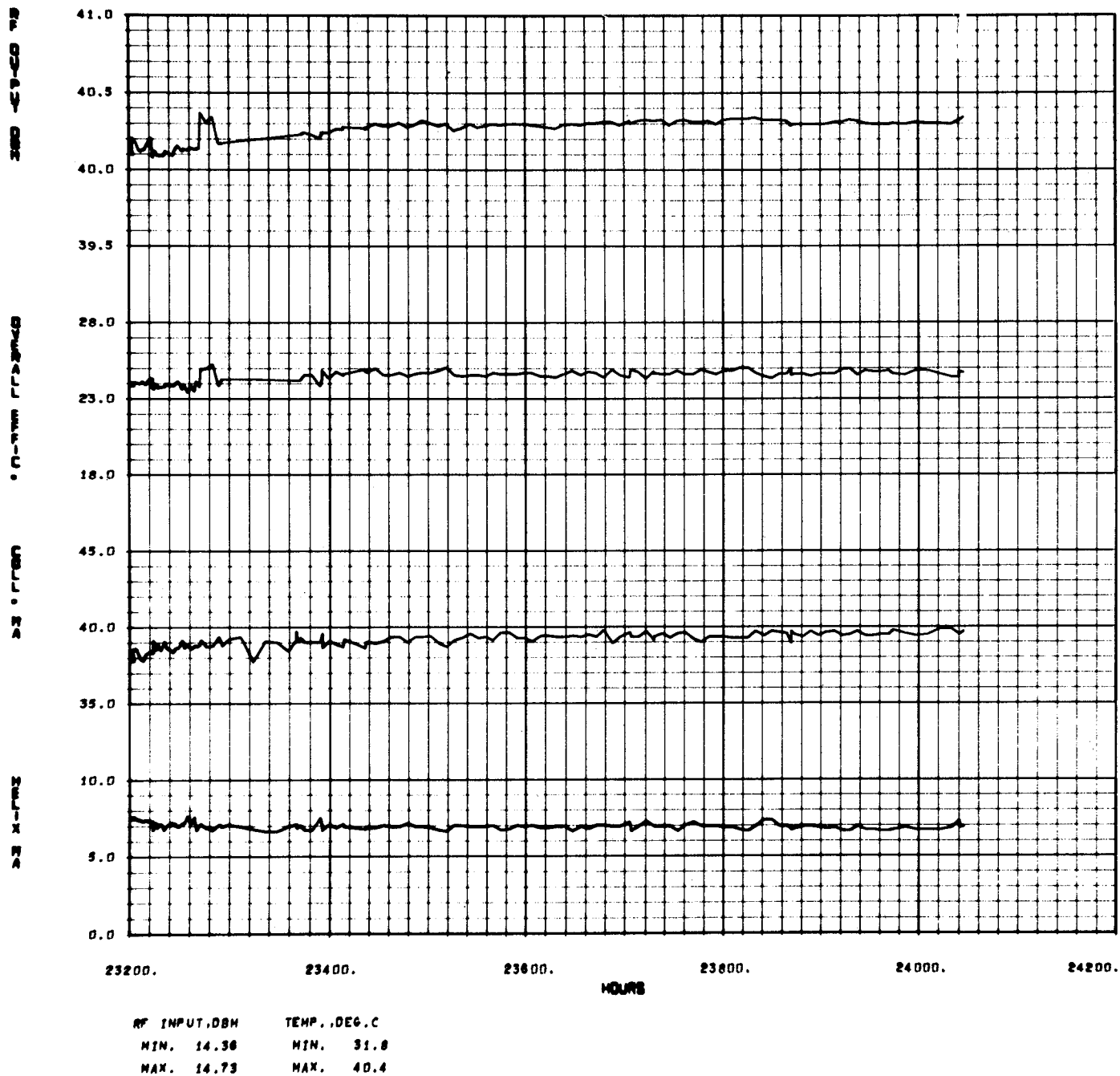


Fig. 32. Computer plot of TWT life test data

E. Low Data Rate Telemetry RF System Development, R. B. Postal

1. Introduction

A solid state MFSK 2295-MHz transmitter is being developed as a subassembly for a telecommunication system capable of surviving a high impact on a planetary surface. A previous article (SPS 37-40, Vol. IV, pp. 198-201) contained design goals, a block diagram and a circuit description of the transmitter. The present status and a description of recent progress are given below.

2. Transmitter Description

Figure 33 is a block diagram of the S-band, high impact, solid state transmitter. The transmitter circuitry is housed in four separate modules to simplify electrical and environmental testing. The MFSK modulator/oscillator circuitry (previously incorporated in module 1) has been placed in a separate module (module 0) to isolate the crystal oscillator from the heat producing amplifiers in module 1.

3. Development Status

The transmitter circuitry and packaging for modules 0, 1, and 2 are being developed at JPL. The crystal assembly used in module 0 is being developed by Valpey Fisher Corporation under JPL contract 951080. The ferrite isolator used in module 3 is being developed by Rantec Corporation under JPL contract 951565. The stripline portion of module 3 was developed by Motorola Corporation.

Engineering models of modules 0, 1, and 2, with the exception of the crystal and a varactor diode, have survived 10,000-g shock levels. The varactor failure occurred at 5200 g's. Additional types of varactors are being evaluated for shock resistance.

Sterilization tests have been performed on modules 1 and 2. These units were subjected to three 14-h temperature cycles of 145°C. No degradation in module performance was measured at the conclusion of these tests; however, there was significant discoloration of the component staking compound. The cause and possible long term effects of the discoloration are being investigated.

Modules 1 and 2 were subjected to RF breakdown tests over a wide range of pressures. No multipactor breakdown was noted at pressures between 10^{-3} and 10^{-6} torr; however, ionic breakdown was observed at four locations in module 2 at a pressure of 1 torr. It is evident that some

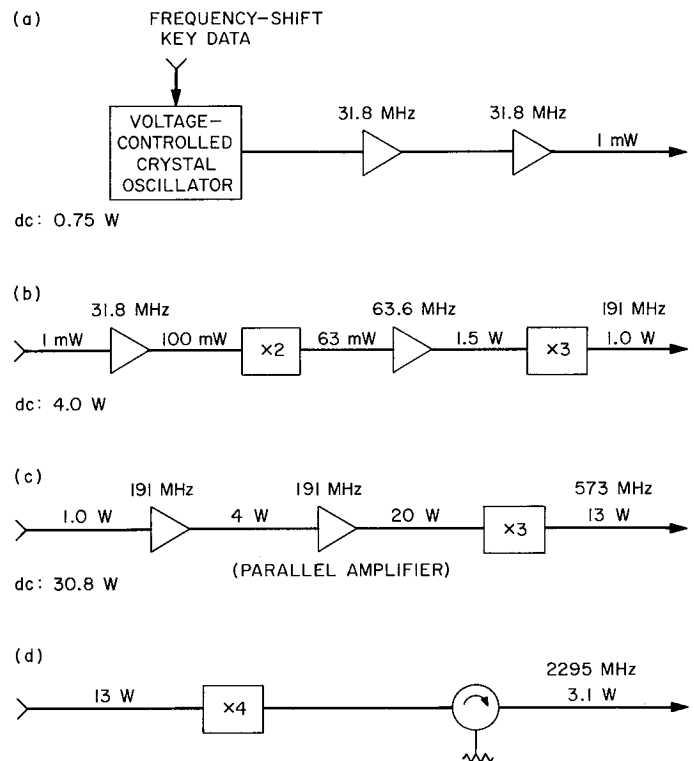


Fig. 33. Transmitter block diagram: (a) module 0; (b) module 1; (c) module 2; (d) module 3

positive form of breakdown suppression (i.e. encapsulation or pressurization in a sealed canister) must be employed in order to insure proper operation in low pressure atmospheres.

Considerable time and effort have been spent in the development of the MFSK modulator for module 0. The modulator circuitry consists of a low gain 31.875 MHz voltage-controlled crystal oscillator followed by two amplifiers. Of prime importance is the oscillator frequency stability necessary to support a low data rate link. The immediate goal is a word separation of only 10 Hz at S-band with a word time of 5 s. This implies a required short term frequency stability ($\Delta f/f$) of the order of $1 \times 10^{-10}/s$. Frequency stability data taken on the bread-board modulator are the following:

Temperature change ($0^\circ - 55^\circ\text{C}$)	$2000 \times 10^{-10}/^\circ\text{C}$
Power supply change ($\pm 1\%$)	6×10^{-10}

The above data show the modulator performance to be unsatisfactory. The frequency change due to power supply variations, however, can be reduced to $0.2 \times 10^{-10}/s$ by improving the power supply regulating circuits. Temperature compensation of the oscillator is extremely difficult

because the Q of the frequency control circuit external to the crystal is very low. The simplest approach appears to be to constrain the frequency change due to temperature by selecting a crystal whose frequency characteristics are $\leq 5 \times 10^{-8}/^{\circ}\text{C}$ and by limiting the rate of temperature change in the oscillator module to $\leq 10^{-3}^{\circ}\text{C/s}$. A preliminary thermal analysis of a passive system (only insulation and heat capacity considered) indicates this low rate is feasible with practical materials. Thus, it appears that the stability goals can be achieved.

Figure 34 is a photograph of the 31.875 MHz, high impact crystal being developed by Valpey Fisher Corporation. The crystal assembly is cylindrical in shape and consists of two plated ceramic holders, a quartz resonator, and a quartz annular ring (not shown). The holder material is 95% pure alumina. The resonator contact surface is gold plated with a chrome undercoat. The solder ring is sintered nickel with a moly-manganese undercoat, and the crystal resonator is a fifth overtone AT-cut blank with gold plated electrodes. The crystal resonator is supported on its periphery by each of the resonator contacts. The holder area within the resonator contact is relieved to a depth of 0.002 in. to allow proper Piezo-electric action of the crystal. Constant periphery pressure is provided by the annular ring, its thickness being the same as that of the resonator. During assembly operations, the resonator and annular ring are clamped between the plated holders, and the complete assembly is then induction solder-sealed in an untreated air atmosphere of $10 \mu\text{m}$.

One such crystal (a 19.125 MHz proof test model unit) was subjected to a series of shock levels through 8600 g 's.

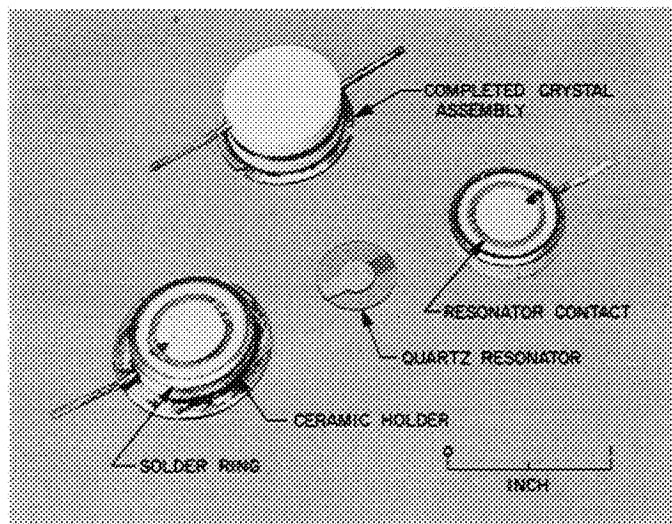


Fig. 34. High impact sterilizable crystal

Test data revealed that frequency shifts due to shock level ranged between 0.05 and 1.5 ppm. The crystal resonator fractured during the third 8600- g shock test. This failure was caused by improper machining of the holder relief cavity.

Valpey Fisher Corporation is presently under contract to produce ten 31.875-MHz crystal units. Delivery of these units is expected December 20, 1967.

The $\times 4$ frequency multiplier (module 3) consists of two stripline doublers laminated onto a central heatsink. The complete unit is shown in Fig. 35. Two doublers are utilized in order to optimize power handling capability and minimize diode junction temperatures. Frequency multiplication at high efficiency is provided by the nonlinear characteristics of the varactor diodes. Figure 36 shows $\times 4$ multiplier output level variations versus temperature when the unit is driven with a 573-MHz level of +41 dBm.

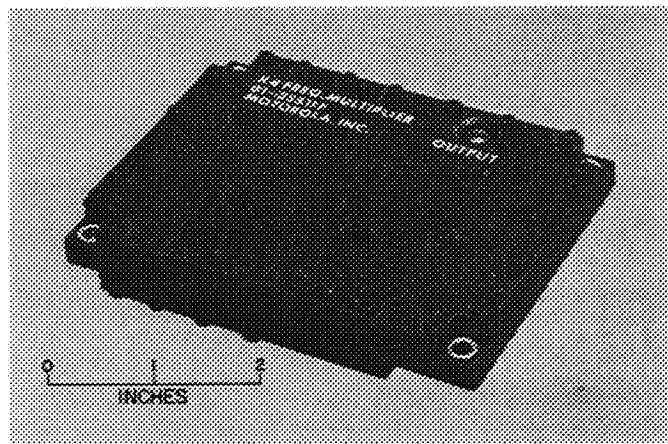


Fig. 35. High impact, sterilizable $\times 4$ frequency multiplier

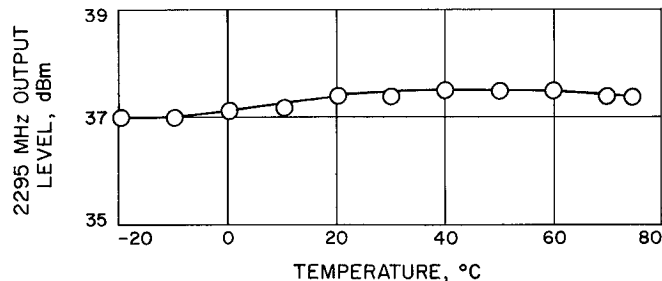


Fig. 36. $\times 4$ frequency multiplier temperature performance

XXIV. Future Projects

ADVANCED STUDIES

A. Lunar Ice, J. R. Bruman and E. C. Auld

1. Introduction

A series of tests to observe the behavior of various mixtures of water and insoluble material as preliminary to a serious attempt to simulate the lunar surface was initially reported in SPS 37-47, Vol. III, pp. 279, 280. These tests continued during this reporting period. The principal purpose of these experiments is to ascertain the longevity of ice deposits beneath the surface layer of dust or sand. A secondary purpose is to seek a plausible explanation for the many visual features of the lunar surface which seem to suggest the action of liquid water.

If the lifetime of buried permafrost were found to be long, the discovery should profoundly affect the course of future lunar exploration. The logistic value of possible lunar water deposits would raise prospecting to a first-priority activity. If any such deposits were reasonably accessible, present limitations on lunar stay time would be removed, and, in addition, the moon would become a base for planetary missions. The payload of a lunar launch, using locally manufactured fuel, would be about 20 times that of a comparable earth-launched mission.

Information concerning the lifetime and behavior of lunar ice should also greatly affect the scientific analysis of the moon's history.

2. Description of Tests and Results

Various frozen samples were subjected to an approximation of solar heating and lunar vacuum in the JPL 6-ft chamber. It should be emphasized that this simulation was applicable not to the arrival of water at the lunar surface, but to its departure; e.g., the suspected deposition of permafrost by plutonic heating implies just the reverse of the present experiments, in which heat was applied from above. Sample behavior was recorded by means of time-lapse photography, embedded thermocouples, and a spring balance supporting the sample.

Sample I, plain ice, experienced peculiar surface activity wherein whisker-like crystals broke off and flew upward after growing to a height of about 2 to 3 mm (Fig. 1). Visual and thermal data indicated direct sublimation with no intervening liquid phase. The rate of loss was approximately 1 mm/h.

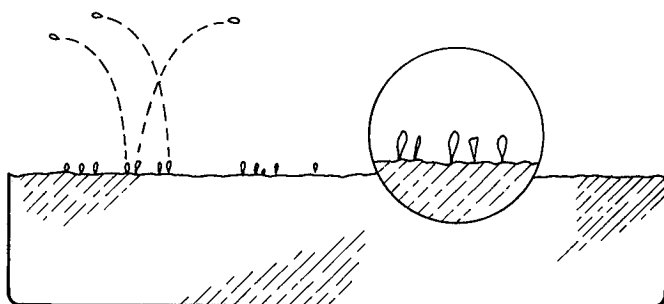


Fig. 1. Form and motion of growths on surface of ice in vacuum (sample I)

Sample II, ice covered with powdered basalt with approximately 0.01-mm particle size to a depth of about 5 mm, also showed mechanical activity. Small eruptions scattered material off the sample, and the surface developed nodular clumps. Ice beneath the basalt powder became covered with small steep craters ranging from 2 to about 10 mm in diameter. No liquid phase was observed. Rate of loss was about 1 mm/h, the same as with sample I. Figure 2(a) shows the nodular surface of the crushed basalt and possible small vent holes. The shape of the ice-basalt interface after the basalt was removed is shown in Fig. 2(b); the apparent melting at the outer edges occurred after removal from the test chamber.

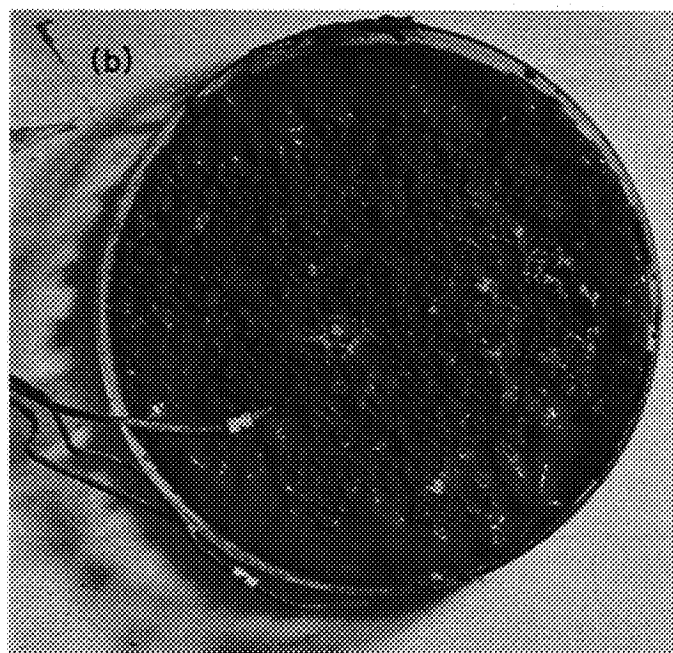
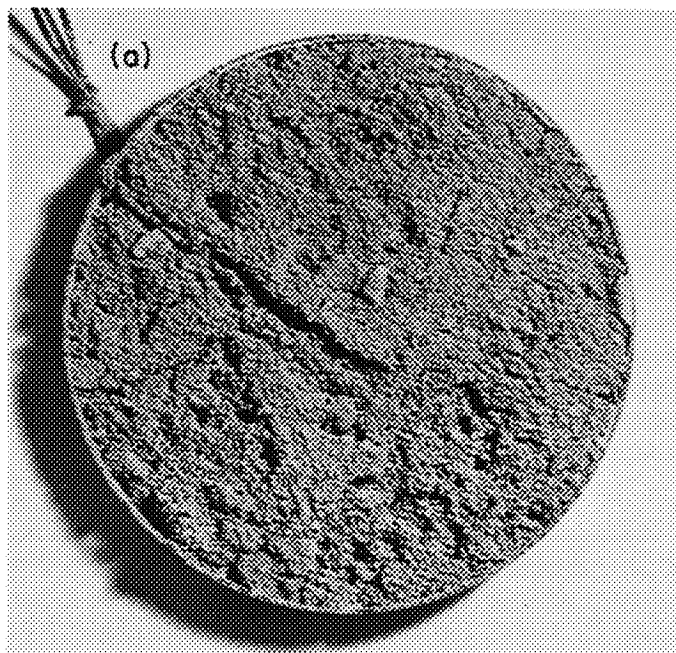


Fig. 2. Sample II: (a) after test, (b) after basalt was removed

Sample III was similar to sample II, except that the basalt powder was 75 mm deep. Liquid water appeared at the interface and formed a layer of mud. Figure 3(a) shows the dark band of liquid at the basalt-ice interface, and Fig. 3(b) shows the frost accretion on the rim of the container and the void above the liquid water in the bottom of the container. Suddenly, apparently after all the ice had melted, the sample erupted violently from its container, leaving about 15 mm of mud in the bottom which immediately froze. In Fig. 3(c), most of the contents have erupted from the container. The two dark bulbous objects in the center are columns of frozen mud near the vertical axis of the container. Small eruptions continued through 1- to 2-mm fumarole-like openings in the surface. Figure 3(d) shows a brief eruption of particles from a vent in the frozen mud left at the bottom of the flask after the main eruption.

3. Conclusions

Apparently the thickness of the dry overburden is a critical factor determining whether loss takes place by continuous sublimation or by intermittent melting and eruption. This will be investigated in future tests. The next series of experiments will use samples of frozen mud covered with a variable depth of dry material.

While not entirely unexpected, the demonstration that liquid water can exist at all under lunar conditions greatly

strengthens hypotheses which stress the importance of lunar water. The additional discovery that such water can, under certain conditions, produce intermittent eruptions suggests that this small-scale simulation may possibly lead to a better understanding of lunar surface

features. It would be premature to suggest that the formation of features such as domes, meandering stream channels, and shrinkage cracks can be meaningfully simulated on a small scale; but, clearly, this possibility should be investigated in future experiments.

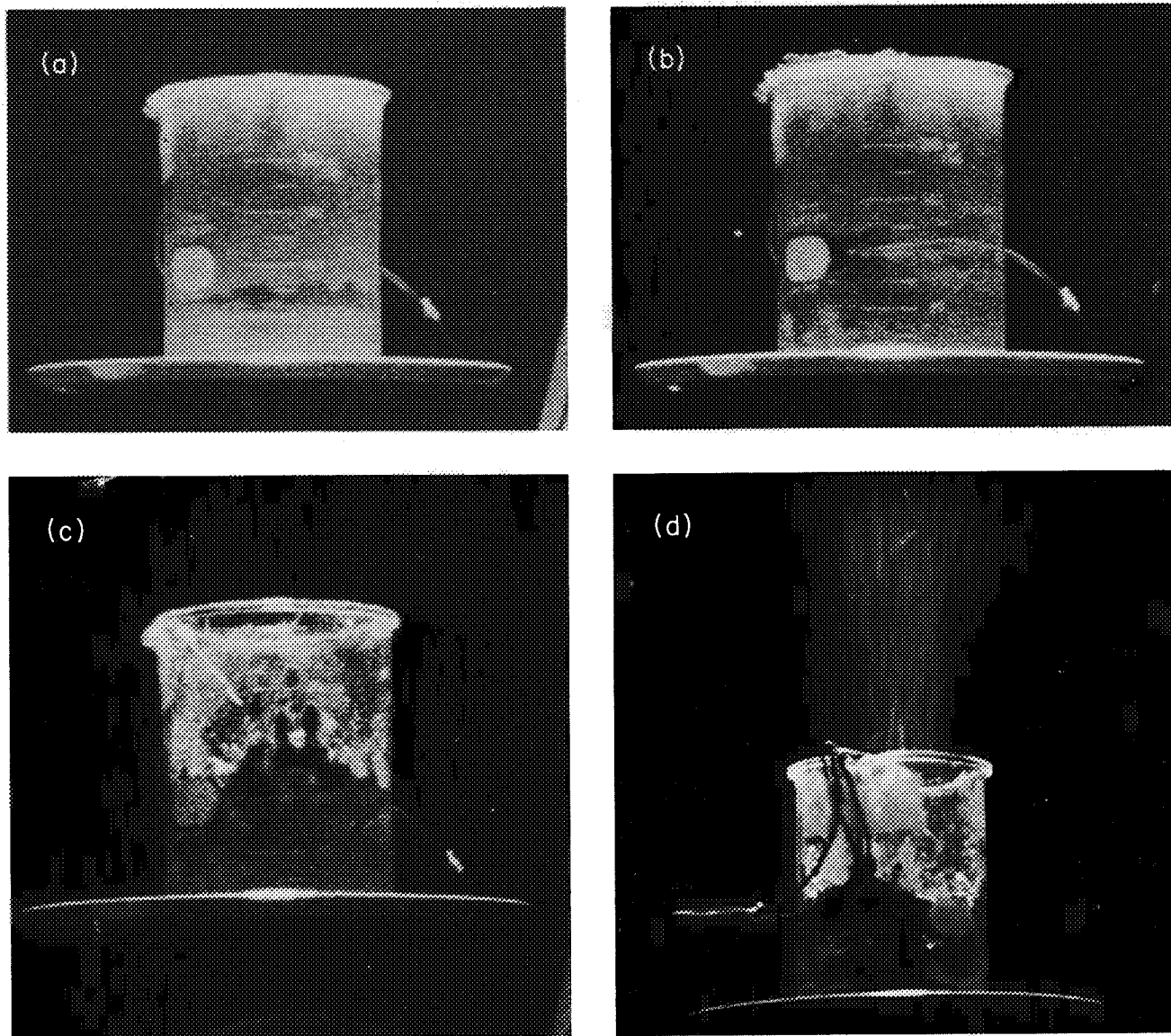


Fig. 3. Sample III: (a) early in test, (b) later, (c) 15 s after (b), (d) during brief eruption after the main eruption

Abbreviations

A/DC	analog-to-digital converter	LS&T	lander sequencer and timer
AEDC	Arnold Engineering Development Center (Tallahoma, Tenn.)	MFSK	multiple frequency-shift-keyed
AM	amplitude modulation	MMT	multiple mission telemetry
ATS	<i>Applications Technology Satellite</i>	MMTD	multiple mission telemetry demodulator
BBB	baseband breadboard	MMTS	multiple mission telemetry system
BCF	beam connection factor	NCO	numerically controlled oscillator
BIBD	balanced incomplete block design	NMR	nuclear magnetic resonance
cdf	cumulative density function	OSE	operational support equipment
CSAD	capsule system advanced development	PCM	pulse-code-modulated
CW	continuous wave	PLOD	planetary orbit determination (program)
DE	development ephemeris	PM	phase-modulated
DSIF	Deep Space Instrumentation Facility	PN	pseudonoise
DSN	Deep Space Network	RBV	return beam vidicon
DSS	deep space station	RC	resistance-capacitance
DVM	digital voltmeter	RRK	Rice-Ramsperger-Kassel
EPD	engineering planning document	SDA	subcarrier demodulator assembly
EPS	entry power subsystem	SDS	Scientific Data Systems
ES&T	entry sequencer and timer	SEC	secondary electron conduction
ETO	ethylene oxide	S/N	signal-to-noise ratio
FM	frequency modulation; feasibility model	SNORE	signal-to-noise ratio estimator
FTS	flight telemetry system	SNR	signal-to-noise ratio
H-P	Hewlett-Packard	SSDPS	solar system data-processing system
IUR	irreducible unitary representation	TIR	total indicator reading
JPL	Jet Propulsion Laboratory	VCO	voltage-controlled oscillator
LPS	lander power subsystem	VSWR	voltage standing-wave ratio
LRC	Lewis Research Center (NASA, Cleveland, Ohio)	YIG	yttrium iron garnet

